The Regulatory Logic of Pax3 Expression in the Neural Tube

Steven James Moore

August 2012

Division of Developmental Biology
MRC National Institute for Medical Research
Mill Hill, London

Department of Cell and Developmental Biology
University College London

Thesis submitted to University College London
for the degree of Doctor of Philosophy

Statement of Declaration

I, Steven James Moore, confirm that the work presented in this thesis is my own, with the exception of figures 3.5 and 3.6, which were kindly provided by Dr Vanessa Ribes. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed:…….. Date:……..

Abstract

The development of complex tissues is dependent on the specification and maintenance of distinct transcriptional identities within populations of equivalent progenitors. In the vertebrate neural tube, progenitor identity is specified by the spatially organised expression of transcription factors within the dorso-ventral (DV) axis of the tissue. A series of selective cross-repressive interactions between transcription factors expressed in adjacent domains forms the basis of a gene regulatory network (GRN), which acts to define and maintain the boundaries between progenitor populations. Pax3 is a key member of the neural tube GRN that is transcribed throughout the dorsal domains of the spinal cord and exhibits a ventral boundary of expression at the level of the sulcus limitans. The regulatory mechanisms that induce Pax3 transcription in neural progenitors and subsequently produce a sharp ventral boundary of gene expression are poorly understood.

Using a genetic lineage tracing strategy in mice, we show that the position of the Pax3 expression domain is established within the neural plate and subsequently maintained throughout spinal cord development. Additionally, we demonstrate that V0 interneuron progenitors transiently express Pax3 during their early development. We employ a comparative genomics based approach to investigate the molecular basis of Pax3 transcription, revealing several novel enhancers associated with the locus. Functional dissection of an early central nervous system (CNS) specific enhancer supports the role of the Wnt signaling pathway in the induction of Pax3 transcription. Furthermore, we establish the function of direct transcriptional repression in the establishment of the Pax3 expression domain. Analysis of a second CNS specific enhancer reveals that Pax3 expression is dependent on autoregulation and Pax7 mediated positive feedback post-neurulation. Together, these data indicate that the temporal activity of two distinct enhancers underlies the regulatory logic of Pax3 expression in the neural tube.

Acknowledgements

I would first like to thank my supervisor, James Briscoe, for the support and advice he has given me throughout my PhD. I feel very fortunate to have had the opportunity to study in his lab. I would also like to thank Vanessa Ribes for her creative input at the start of this project and the continued collaboration throughout. I am particularly grateful to her for allowing me to include unpublished data in this thesis. I would like to thank all members of the Briscoe lab, past and present, for their insightful discussions that have greatly contributed to this work. I am grateful for the tools provided by Fredéric Relaix, Philip Ingham, Claudia Seger, Javier Terriente, Nipam Patel and Simon Hughes that were vital for the completion of this project. Finally, I would like to thank Toni and my family whose support made this thesis possible.

Contents

List of figures

List of Tables

Abbreviations

^{32}P	=	Phosphorus-32
AP	=	Anterior-posterior
BAC	=	Bacterial artificial chromosome
bHLH	=	Basic helix-loop-helix
BLAST	=	Basic local alignment search tool
BMP	=	Bone morphogenetic protein
bp	=	Base pair
ChAT	=	Choline acetyltransferase
ChIP	=	Chromatin immunoprecipitation
ChIP-chip	=	ChIP followed by microarray
ChIP-seq	=	ChIP followed by high-throughput sequencing
CNE	=	Conserved non-coding element
CNS	=	Central nervous system
CRM	=	Cis-regulatory module
dP	=	Dorsal progenitor
DV	=	Dorso-ventral
E	=	Embryonic day
EC	=	Embryonal carcinoma
ECR	=	Evolutionary conserved region
EMSA	=	Electrophoretic mobility shift assay
ESC	=	Embryonic stem cells
FGF	=	Fibroblast growth factor
FP	=	Floor plate
Fugu	=	*Takifugu Rubripes*
GFP	=	Green fluorescent protein
GRN	=	Gene regulatory network
H3K27ac	=	Acetylation of histone H3 at lysine 27
H3K27Me3	=	Trimethylation of histone H3 at lysine 27
H3K4Me1	=	Monomethylation of histone H3 at lysine 4
H3K4Me3	=	Trimethylation of histone H3 at lysine 4
HD	=	Homeodomain
HH	=	Hamburger and Hamilton stage
hpe	=	Hours post-electroporation

hpf	=	Hours post-fertilisation
kb	=	Kilobases
LB	=	Luria-Bertani
Lef	=	Lymphoid enhancer factor
MiniTol2-Citrine	=	MiniTol2 TKProm Gal4-UAS Citrine expression vector
NCE	=	Neural crest enhancers
p	=	Progenitor domain
P	=	Postnatal day
Pax	=	Paired box containing
Pax3/7	=	Pax3 and Pax7
PBS	=	Phosphate buffered saline
PCR	=	Polymerase chain reaction
PD	=	Paired domain
PFA	=	Paraformaldehyde
PSPM	=	Position specific probability matrix
Ptc	=	Patched
RA	=	Retinoic acid
RBPJ	=	recombination signal binding protein for immunoglobulin kappa J
RP	=	Roof plate
Shh	=	Sonic hedgehog
Sim	=	*Single-minded*
Smo	=	Smoothened
Tcf	=	T-cell factor
TFBS	=	Transcription factor binding sites
YFP	=	Yellow fluorescent protein

Chapter 1: Introduction

1.1 – The formation and patterning of the neural tube

The spinal cord is an anatomically and functionally specialised tissue that enables vertebrates to perceive multiple forms of stimuli, relay this information to higher brain levels and execute the appropriate motor responses. Sensory interneurons contained within the grey matter of the dorsal horn receive inputs from primary afferent neurons via the dorsal root ganglion. These signals can either be directly relayed to motor neurons within the ventral horn in the form of a spinal reflex or conveyed to higher processing centres via ascending white matter tracts. Motor commands from the brain are subsequently relayed to the appropriate level of the ventral spinal cord via descending tracts in order to coordinate the activity of specific muscle groups (Goulding et al., 2002). Centuries of anatomical and physiological studies have resulted in a wealth of knowledge describing the spinal cord in great detail. By comparison, our understanding of the cellular and molecular mechanisms that act to define and subsequently maintain molecularly unique territories within the developing tissue is less complete.

1.1.1 – The formation of the embryonic central nervous system

The presumptive CNS is induced within the dorsal ectoderm of the blastula stage embryo by the activity of bone morphogenetic protein (BMP) pathway inhibitors (Stern, 2005). Following gastrulation, these cells reside dorsally along the length of the embryo as an epithelial sheet, known as the neural plate. Neural progenitors are initially specified to an anterior identity and are subsequently posteriorised by the graded activity of the Wnt, fibroblast growth factor (FGF) and retinoic acid (RA) signaling pathways in the caudal regions of the neural plate (Durston et al., 1989 ; Kiecker and Niehrs, 2001 ; Nordstrom et al., 2002 ; Storey et al., 1998). Anterior neuronal identities are maintained by the activity of secreted pathway antagonists, revealed by the loss of anterior structures in embryos lacking these factors or by misexpression of posteriorising signals across the anterior-posterior (AP) axis (Böttcher and Niehrs, 2005 ; Maden, 2007 ; Niehrs, 2004).

Following AP patterning, the neural plate is transformed by a stereotypical series of morphogenetic movements during neurulation. This process is initiated by the

bending of the lateral edges of the neural plate at the point of contact with the non-neural ectoderm, giving rise to the neural folds. The lateral folds of the neural plate elevate towards to the midline where they fuse and subsequently segregate from the overlying non-neural epithelium to form the dorsal midline of the neural tube (Copp et al., 2003). This sequence of morphogenetic movements is initiated at the boundary between the presumptive hindbrain and spinal cord, before proceeding anteriorly to give rise to the brain and its respective subdivisions and posteriorly to form the embryonic spinal cord (Copp et al., 2003).

The developing neural tube is one of the best-studied examples of developmental patterning as it represents an experimentally tractable model in which a population of genetically equivalent progenitors give rise to a stereotypic distribution of neuronal subtypes across the DV axis of the tissue (Jessell, 2000). The process of DV patterning is initiated by the graded activity of several morphogens secreted from the roof plate (RP) and floor plate (FP) organising centres within the dorsal and ventral midlines of the tissue, respectively (Fig. 1.1). Morphogen signaling acts to establish a spatially organised transcription factor code consisting of homeodomain (HD) and basic helix-loop-helix (bHLH) containing proteins within the ventricular zone of the neural tube (Lewis, 2006) (Fig. 1.1). The patterned distribution of transcription factors across the DV axis of the neuroepithelium creates 11 molecular distinct progenitor domains (p) within the tissue, which directly correlate with the neuronal subtype a given progenitor is fated to form during neurogenesis (Fig. 1.1) (Briscoe et al., 2000). Each of the 6 dorsal progenitor domains (dP) within the neural tube gives rise to the sensory interneurons of the spinal cord. Within the ventral neural tube, pMN progenitors give rise to motorneurons whereas the remaining 4 domains give rise to ventral classes of interneurons (Fig. 1.1). In the following sections, I will outline the morphogen signaling pathways that act to pattern the neural tube and discuss the key studies that provided insight into their function.

Copyright restricted material has been removed from this digital copy

Fig. 1.1 – The dorso-ventral patterning of the vertebrate neural tube

Progenitors within the neural tube are subject to Shh signaling from the FP and BMP/Wnt signaling from the RP. The combined activity of morphogens in the neural tube establishes 11 molecularly distinct progenitor domains across the DV axis of the tissue. The transcriptional identity of progenitors directly correlates with the subtype of neurons they are fated to give rise to during the process of neurogenesis.

Figure adapted from Lewis, 2006.

1.1.2 – Sonic hedgehog signaling patterns the ventral neural tube

Sonic hedgehog (Shh) ligand is initially produced by the notochord, a rod-like structure of mesodermal cells that underlies the floor plate. Signaling from the notochord induces a second source of Shh secretion within the FP of the ventral neural tube (Marti et al., 1995). The secretion of Shh from the FP creates a gradient of ligand throughout the ventricular zone of the neural tube, resulting in a high concentration ventrally that decreases as it spreads dorsally (Fig. 1.1) (Chamberlain et al., 2008). The Shh pathway functions by modifying the activity of the Gli family transcription factors, which are processed to repressive forms in the absence of ligand. Shh binding to its receptor patched (Ptc) results in the de-repression of the transmembrane protein smoothened (Smo). Once active, Smo signals intracellularly by promoting the processing of the Gli transcription factors to activator forms, which enter the nucleus and induce the transcription of Shh target genes (Ingham and McMahon, 2001).

The cellular interpretation of the Shh gradient in the neural tube was shown to be dependent on both the concentration of ligand and the duration of signaling, due to temporal adaption within the pathway (Dessaud et al., 2007). In this model, cells are initially hypersensitive to Shh and consequently high levels of positive Gli transcriptional activity are produced in response to a low concentration of the ligand. Over time, cells become less responsive to a stable concentration of Shh via a negative feedback loop, in which cells receiving Shh signaling induce the expression of Ptc. The effect of this pathway mediated feedback is the desensitisation of the cell to Shh by increasing the number of receptors available for ligand binding, consequently reducing the Gli activity achieved by a constant level of Shh over time. In order to maintain a high level of Gli activity, cells must receive a high concentration of Shh over time to compensate for the increase in Ptc expression that pathway activity induces (Dessaud et al., 2007).

Shh signaling in the ventral neural tube acts to repress genes expressed dorsally and induce the transcription of ventral markers; commonly referred to as class I and class II genes, respectively (Briscoe and Ericson, 2001). The morphogen activity of Shh was determined in a series of experiments in which explanted neural plate tissue was cultured in the presence of exogenous ligand. These data revealed that class I genes are repressed by increasing concentrations of Shh at a threshold that is indicative of their expression domain, for example Pax7 was shown to be repressed at a lower concentration than Irx3 (Fig. 1.1) (Briscoe et al., 2000 ; Ericson et al., 1996).

Conversely, the class II genes Nkx6.1 and Nkx2.2 were induced in a concentration dependent manner according to their expression profile in the ventral neural tube (Briscoe et al., 2000). Loss of Shh signaling in the neural tube is characterised by the de-repression of class I genes and a failure of class II gene induction (Briscoe and Ericson, 2001 ; Chiang et al., 1996 ; Ericson et al., 1996 ; Wijgerde et al., 2002). Conversely, ectopic expression of Shh, Smo or Gli has been shown to be sufficient to induce the expression of class II genes throughout the neural tube at the expense of class I genes (Ericson et al., 1996 ; Hynes et al., 2000 ; Stamataki et al., 2005 ; Ribes et al., 2010). Taken together, these studies demonstrate the requirement of graded Shh signaling for the correct assignment of progenitor identity in the ventral neural tube and the repression of dorsally expressed genes.

1.1.3 – BMP and Wnt signaling coordinate dorsal patterning

Experiments performed in chick embryos revealed the instructive role of morphogens secreted from the RP by demonstrating that this tissue was sufficient to induce the expression of dorsal interneuron markers in neural plate explants (Liem et al., 1997). Subsequent studies in transgenic mice confirmed the requirement for RP by genetically targeted expression of diptheria toxin, revealing the respecification of dI1 and dI2 neurons to a dI3 identity (Lee et al., 2000). These data established the RP as the second organising centre of the neural tube, whose activity was required for the correct assignment of dorsal neuronal identities.

The patterning activity of the RP is achieved, in part, by the secretion of several members of the bone morphogenetic protein family (Lee et al., 2000 ; Liem et al., 1997). BMP signaling is activated by ligand binding to a heterodimer of type I and type II receptors, which in turn phosphorylate Smad1/5/8, the downstream effectors of the pathway. Phosphorylated Smad1/5/8 form complexes with Smad4 that translocate to the nucleus to regulate gene expression (Massague et al., 2005). The role of BMP signaling in DV patterning was established in experiments that demonstrated the induction of dorsal interneuron markers in neural plate explants cultured with exogenous ligand (Liem et al., 1997). Expression of activated forms of BMP receptors 1A or 1B throughout the chick neural tube revealed that BMP signaling is sufficient to induce the expression of class I genes in the ventral spinal cord and consequently alter the fate of these cells (Timmer et al., 2002). In the converse experiment, inhibition of BMP signaling using the pathway antagonist noggin or RNAi targeted to Smad4 revealed the

loss of dI1 interneurons and a compensatory dorsal shift of the dI2-4 populations towards the RP (Chesnutt et al., 2004).

In addition to BMPs, the RP also secretes members of the Wnt family of morphogens (Hollyday et al., 1995 ; Parr et al., 1993). The Wnt signaling pathway regulates target genes by modulating the stability of its transcriptional effector, β-catenin. In the absence of ligand, intracellular β-catenin is degraded by the activity of the destruction complex. Wnt binding to its receptor complex acts to stablise β-catenin within the cell by sequestering key components of the destruction complex at the membrane. Once stabilised, β-catenin translocates to the nucleus where it acts in concert with T-cell factor (Tcf)/ lymphoid enhancer factor (Lef) family transcription factors to regulate gene expression (Logan and Nusse, 2004). Mice lacking both *Wnt1* and *Wnt3a* revealed a loss of dI1 and dI2 markers and a compensatory increase in dI3 neurons, a phenotype comparable to that observed in RP ablation experiments (Lee et al., 2000 ; Muroyama et al., 2002). Moreover, loss of both *Wnt1* and *Wnt3a* did not affect the expression of BMP ligands in the roof plate, suggesting that the Wnt pathway is epistatic to BMP in the dorsal neural tube (Muroyama et al., 2002). Ectopic expression of Wnt ligands or activated β-catenin was shown to result in the dorsalisation of the neural tube, assessed by both progenitor identity and neuronal subtype (Alvarez Medina et al., 2008 ; Muroyama et al., 2002 ; Zechner et al., 2007). Together, these studies demonstrate the function of RP derived morphogens in assigning positional identity within the dorsal spinal cord and the repression of ventral fates.

1.1.4 – The spinal cord gene regulatory network

The combined activity of morphogen signaling in the ventricular zone of the neural tube results in the establishment of 11 progenitor domains along the DV axis, each of which expresses a distinct complement of HD and bHLH transcription factors (Fig. 1.1) (Briscoe et al., 2000). This transcription factor code forms the basis of the neural tube GRN, which functions to establish and maintain gene expression boundaries between adjacent domains by cross repression (Dessaud et al., 2008). One of the best described repressive circuits within the neural tube GRN acts to delineate the p0, p1 and p2 domains of interneuron progenitors. The p2 marker Nkx6.1 was shown to repress the p1 markers Nkx6.2 and Dbx2 in both gain and loss of function studies (Fig. 1.2A) (Briscoe et al., 2000 ; Sander et al., 2000 ; Vallstedt et al., 2001). This interaction was shown to be mutually cross-repressive in the case of Dbx1 by gain of function in the

chick neural tube (Fig. 1.2A) (Briscoe et al., 2000). The ventral boundary of the p0 domain was shown to be dependent upon the repression of Dbx1 by Nkx6.2, expressed in the adjacent p1 domain (Fig. 1.2A) (Vallstedt et al., 2001). Thus, these data demonstrate how a series of selective repressive interactions between 4 transcription factors are sufficient to define and maintain the boundaries of 3 adjacent interneuron progenitor domains.

The boundary of the p3 and pMN domains of the ventral neural tube are established and maintained by a similar repressive mechanism involving the transcription factors Nkx2.2, Olig2 and Pax6. Ectopic expression of the p3 marker Nkx2.2 by chick *in-ovo* electroporation was shown to repress both Pax6 and Olig2 (Fig. 1.2B) (Briscoe et al., 2000 ; Novitch et al., 2001). Pax6, a class II gene expressed at low levels in the pMN domain, was shown to repress Nkx2.2 in loss of function studies, establishing a mutually cross-repressive interaction (Fig. 1.2B) (Ericson et al., 1997). Gain and loss of function of the pMN determinant Olig2 revealed repressive interactions with both Pax6 and Nkx2.2 (Balaskas et al., 2012). Together, this network of repressive interactions at the p3/pMN boundary acts to delineate the progenitors of V3 interneurons from motor neurons (Fig. 1.2B).

Recent work has highlighted the importance of the GRN and demonstrated that the Shh response in the ventral neural tube is an emergent property of the network architecture (Balaskas et al., 2012). These data demonstrate that at the onset of patterning, low levels of pathway activation cannot induce the expression of either Olig2 or Nkx2.2 and consequently Pax6 is expressed throughout the ventral neural tube. The concentration of extracellular Shh increases over time, raising the level of positive Gli transcriptional activity intracellularly within the ventral spinal cord (Balaskas et al., 2012 ; Chamberlain et al., 2008). Olig2 transcription is induced at a moderate level of Gli activity, resulting in the repression of Pax6 ventrally (Fig. 1.2B) (Balaskas et al., 2012). Nkx2.2 cannot be induced by moderate levels of Gli transcriptional activity, due to the repressive interactions of both Olig2 and Pax6 (Fig. 1.2B) (Balaskas et al., 2012). The activation of Nkx2.2 transcription is achieved at the maxima of Gli activity, resulting in a switch in the GRN by its repression of both Olig2 and Pax6 (Fig. 1.2B) (Balaskas et al., 2012). The level of extracellular Shh ligand continues to increase in the neural tube during development, however intracellular Gli activity decreases to a level below that required for Nkx2.2 activation, due to temporal adaption within the pathway (Balaskas et al., 2012 ; Chamberlain et al., 2008 ; Dessaud et al., 2007). Despite the reduction in intracellular pathway activity, the expression profiles of genes within the

ventral neural tube remain stable (Balaskas et al., 2012). This robustness was shown to be due to the ability of Nkx2.2 to repress both Olig2 and Pax6, meaning lower levels of Gli activity are required to maintain Nkx2.2 expression than induce its transcription within a repressive network (Balaskas et al., 2012). Thus, the architecture of the GRN in the ventral neural tube provides a mechanism for translating a non-linear gradient of intracellular shh pathway activity to bistable expression states and confers robustness to neural patterning.

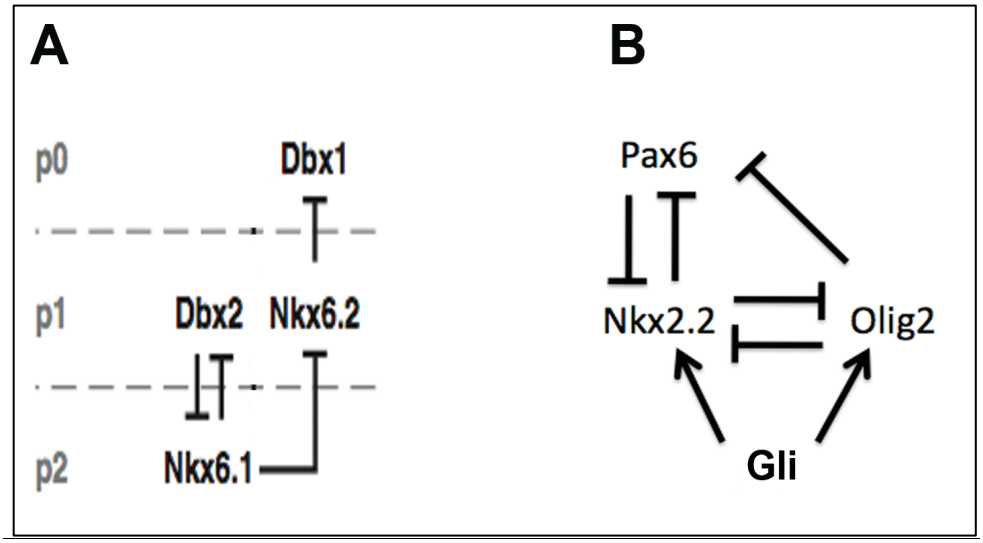


Fig. 1.2 – The ventral neural tube GRN is based on selective cross-repression

The boundaries between the p2, p1 and p0 domains of interneuron progenitors are defined and maintained by a series of repressive interactions between Nkx6.1, Nkx6.2, Dbx1 and Dbx2 (A). Similarly, the boundary between the p3 and pMN domains is defined and maintained by the repressive interactions between Nkx2.2, Olig2 and Pax6 (B). Recent studies have shown that the cellular interpretation of Shh signaling in the neural tube is an emergent property of this repressive network (Balaskas et al., 2012).

Figure 1.2A adapted from Dessaud et al., 2008.

1.2 – The molecular basis of gene regulation

Tissue patterning during embryonic development is achieved by the highly coordinated spatiotemporal expression profiles of transcription factors within a population of progenitors, which acts to impart cells with a unique molecular identity and consequently determines their fate. Mechanistically, the expression profile of a given gene over time is the result of a series of interactions between the non-coding cis-regulatory modules (CRMs) associated with the locus. Thus, considerable research effort in the field of developmental biology is devoted to the elucidating the molecular interactions that underlie the regulatory logic of tissue patterning networks. Central to this process is the core promoter, a specialized CRM located immediately upstream of the transcriptional start site of a gene. This region provides a docking site for the core transcriptional machinery, permitting the binding of RNA polymerase II and the initiation of transcription (Fig. 1.3A). The transcriptional activity of promoters is coordinated by enhancers. These CRMs provide a platform of transcription factor binding sites (TFBS) that determine the expression profile of a given gene over time by directly contacting the promoter and initiating transcription (Fig. 1.3B). Finally, the specificity of enhancer-promoter interactions is determined by insulators, which function to restrict enhancer binding to specific domains within the genome (Fig. 1.3C). In the following section, I will discuss the role of transcriptional enhancers in defining the spatiotemporal expression profile of developmental genes and highlight the properties of these CRMs that facilitate their identification within the genome.

Copyright restricted material has been removed from this digital copy

Fig. 1.3 – The mechanistic basis of differential gene regulation

Proximal promoters recruit the general transcriptional machinery to the transcription start sites of genes (A). Enhancers are a distal class of CRM that provide binding sites for cell-type specific transcription factors. Once active, enhancers physically interact with promoters to coordinate their spatiotemporal activity (B). Insulators regulate the specificity of enhancer – promoter interactions by limiting enhancer translocation to defined regions within the genome (C).

Figure adapted from Noonan and McCallion, 2010.

1.2.1 – Enhancer activity is the basis of differential gene expression

Transcriptional enhancers were first described in studies investigating the regulatory function of a 72 base pair (bp) repeat within the SV40 viral genome (Banerji et al., 1981 ; Moreau et al., 1981). Cells transfected with recombinant plasmids harbouring this element were shown to exhibit a significantly increased level of transcription over controls, regardless of its orientation or position to either an endogenous or heterologous promoter (Banerji et al., 1981 ; Moreau et al., 1981). Subsequent studies performed within metazoan genomes revealed that enhancers could be located within the introns of their associated genes, as demonstrated for the murine immunoglobulin heavy chain locus (Banerji et al., 1983 ; Gillies et al., 1983). These investigations were also the first to demonstrate that enhancers possess the ability to regulate transcription in a cell type specific manner, by assaying the activity of the immunoglobulin enhancer in lymphocytes and fibroblasts (Banerji et al., 1983 ; Gillies et al., 1983). These data decoupled the spatiotemporal regulation of transcription from the core promoter and established distal enhancers as the basis of differential gene expression.

The initial observation that enhancers could be located distally to their endogenous promoter was shown to be a common property of these CRMs in subsequent investigations. This feature is highlighted by a series of studies assessing the regulation of *Shh* transcription in the developing limb bud, which was shown to be dependent on an enhancer located over 1 megabase away from its promoter (Lettice et al., 2002 ; Lettice et al., 2003). Mutation of this enhancer was shown to be the underlying cause of the human congenital limb malformation preaxial polydactyly (Lettice et al., 2003). Transgenic mice lacking this CRM exhibit a complete loss of *Shh* expression in the limb bud, resulting in severe limb truncations (Sagai et al., 2005). These data demonstrate the essential roles that enhancers perform in the coordination of tissue patterning and their emerging roles in human disease. However, they also raise important questions concerning the mechanism of enhancer-promoter interactions over large genomic intervals.

The prevailing model of gene regulation mediated by distal enhancers is commonly termed ‘looping’ (Ptashne, 1986). In this model, the expression profile of a given gene is determined by the physical association of a cell type specific enhancer with the promoter, achieved by the rearrangement or ‘looping out’ of the intervening DNA (de Laat et al., 2008 ; Ptashne, 1986). This hypothesis has been extensively

supported by investigations utilising novel techniques designed to assess the 3 dimensional structure of chromatin, such as the chromosome conformation capture assay and its subsequent variations (Dekker et al., 2002 ; Miele and Dekker, 2008). The first of these studies to validate the looping hypothesis *in-vivo* demonstrated that regulatory modules associated with the murine *β-globin* locus were physically associated with the promoter in erythroid cells, in which the gene cluster is expressed. Conversely, this interaction was not observed in non-expressing brain tissue (Tolhuis et al., 2002). This mechanism of enhancer-promoter interaction was subsequently demonstrated to be a general principle of transcriptional regulation, validating the looping model as the basis of differential gene expression (Amano et al., 2009 ; Miele and Dekker, 2008 ; Splinter and de Laat, 2011).

Despite significant advances in the functional description of enhancers, the *de novo* identification of these regulatory elements is complicated by their ability to function from any position in the locus. Consequently, studies attempting to identify enhancers were frequently based on the serial deletion of a large genomic fragment with an established regulatory function. Putative minimal enhancers were subsequently assayed for regulatory potential by their ability to activate the transcription of a reporter gene when placed proximally to a minimally active promoter (Epstein et al., 1999 ; Sasaki and Hogan, 1996). Although this is a reliable method of enhancer identification, it is both labour intensive and gene centric, prohibiting the genome wide identification of putative regulatory regions.

1.2.2 – Comparative genomics as a method of enhancer discovery

The first study to suggest that the conservation of non-coding sequence might be predictive for the identification of functional CRMs was based upon the hypothesis that enhancers, like coding sequence, were maintained under negative selective pressure during evolution (Aparicio et al., 1995). This theory was tested by comparing the sequence of the Hoxb-4 locus in the mouse and *Takifugu Rubripes* (fugu) genomes, which last shared a common ancestor approximately 430 million years ago (Aparicio et al., 1995). These data revealed the presence of a 92 bp conserved non-coding element (CNE) located within an intron and two CNEs within the 3' region of the locus (Aparicio et al., 1995). These CNEs resided in regions of the locus that had previously been shown to contain the regulatory elements that coordinated *Hoxb-4* expression in mice (Whiting et al., 1991). Analysis of the activity mediated by the defined fugu CNEs

demonstrated that these regions were both necessary for the function of previously defined enhancers and sufficient to direct transgene expression within the endogenous *Hoxb-4* domain of mouse embryos. (Aparicio et al., 1995). Together, these data established the conservation of non-coding sequence as a predictive feature of enhancers, regardless of their position within the genome. Furthermore, this report was the first to demonstrate the value of including the fugu genome in this type of comparative genomic study.

The publication of several model genomes facilitated the application of comparative genomics as a high-throughput method of *de-novo* enhancer discovery. One of the first studies to assess the genome wide conservation of non-coding sequence identified 481 mammalian ultraconserved elements, defined as being 100% homologous over at least 200 bp in the human, mouse and rat genomes (Bejerano et al., 2004). Ultraconserved regions were shown to be enriched within the loci of transcription factors or developmental regulators in the human genome, suggesting that they may function as enhancers, however the *in-vivo* function of these regions was not assessed during this study (Bejerano et al., 2004).

The publication of the fugu genome permitted the genome wide assessment of non-coding sequence conservation over large spans of vertebrate evolution (Aparicio et al., 2002). The first study to utilise this evolutionary distance to predict functional regulatory elements was based upon pairwise alignments of the human and fugu genomes, revealing 1,373 CNEs of at least 100 bp in length (Woolfe et al., 2005). The authors noted that these CNEs were conserved with an average identity of 84%, indicating that the integrity of these regions were maintained under a negative selection comparable to that observed within coding exons (Woolfe et al., 2005). Analysis of the location of each CNE within the human genome revealed that these 1,373 regions were not randomly distributed and instead clustered into 165 groups, each containing multiple CNEs. 93% of CNE clusters were shown to be associated with the loci of key regulators of embryonic development (Woolfe et al., 2005). The regulatory potential of defined CNEs was assessed in zebrafish embryos, a relatively high-throughput method previously validated for this type of study (Müller et al., 1997 ; Müller et al., 1999 ; Woolfe et al., 2005). These data revealed that 92% of the selected CNEs functioned as enhancers *in-vivo*, capable of directing transcription in a profile that frequently recapitulated the expression domain of their associated genes (Woolfe et al., 2005).

The predictive power of comparative genomics was further demonstrated by a study that compared the *in-vivo* activity of human-fugu CNEs and mammalian

ultraconserved regions (Pennacchio et al., 2006). 137 human-fugu CNEs with at least 70% identity between the two genomes were assessed for transcriptional activity in mouse embryos at embryonic day (E) 11.5, 42% of which were shown to function as tissue specific enhancers (Pennacchio et al., 2006). The CNE dataset was compared with the activity of 84 mammalian ultraconserved elements, of which 60% were shown to act as enhancers (Pennacchio et al., 2006). The authors demonstrated that these two datasets significantly overlapped, such that 39% of the human-fugu CNEs assessed were also perfectly conserved over at least 200 bp in mammalian genomes. 61% of human-fugu-ultraconserved elements were shown to be tissue specific enhancers at E11.5 in mice, a figure that is highly likely to underestimate the regulatory potential of these regions by assaying at only one developmental stage (Pennacchio et al., 2006).

These studies established the conservation of enhancers as a fundamental property of genome evolution, which acts to ensure the correct regulation of developmental genes during embryogenesis. This hypothesis is supported by a number of mutations within conserved enhancers that have been shown to disrupt gene expression, resulting in human disease and developmental defects (Kleinjan and Van Heynigen, 2005). In contrast to these findings, however, a study has demonstrated that a selected set of ultraconserved enhancers are not required for the correct regulation of key developmental genes during mouse embryogenesis (Ahituv et al., 2007). Targeted deletions of elements associated with the *Pax6*, *Arx*, *Sox3* and *Dmrt1/ Dmrt3* loci resulted in the production of viable mice that did not exhibit any of the severe developmental defects associated with the mutation of these genes (Ahituv et al., 2007). These data demonstrate that the degree to which a regulatory element is conserved is not a reliable predictor of the requirement for its activity *in-vivo*. Furthermore, these results suggest that certain loci might harbour several functionally redundant enhancers, which act to provide robustness to developmental transcription.

1.2.3 – Post-genomic methods of enhancer identification

The development of high-throughput techniques, such as chromatin immunoprecipitation (ChIP) followed by microarray (ChIP-chip) or sequencing (ChIP-seq), facilitated the annotation of CRMs on a genome wide scale based on their chromatin signatures. The survey of 1% of the human genome performed by the ENCODE consortium revealed that active promoters are associated with high levels of trimethylation of histone H3 at lysine 4 (H3K4Me3) (The ENCODE project consortium,

2007). In contrast, enhancers are marked by monomethylation of histone H3 at lysine 4 (H3K4Me1), a feature that was shown to be predictive of their regulatory function (The ENCODE project consortium, 2007 ; Heintzman et al., 2007).

In addition to the methylation status of lysine 4, the acetylation of histone H3 lysine 27 has been shown to be predictive of transcriptionally active enhancers (Creyghton et al., 2010 ; Rada-Iglesias et al., 2011). This chromatin signature of enhancer activation was established by assessing the histone modifications present on enhancers within human embryonic stem cells (ESC), revealing two distinct classes (Rada-Iglesias et al., 2011). Enhancers associated with genes actively transcribed in ESC were marked by H3K4Me1 and acetylated histone H3 lysine 27 (H3K27ac) (Rada-Iglesias et al., 2011). In contrast, non-transcribed developmental genes were marked by H3K4Me1 and the polycomb mediated repressive signature, histone H3 lysine 27 trimethylation (H3K27Me3) (Rada-Iglesias et al., 2011 ; Spivakov and Fisher, 2007). Upon ESC differentiation to a neuronal progenitor fate, the chromatin status of developmental enhancers was observed to be enriched for K3K4Me1 and H3K27ac in the absence of H3K4Me3, suggesting that the temporal activity of these regulatory elements is reflected by their histone code (Rada-Iglesias et al., 2011). Furthermore, the activity of developmental enhancers, identified by their poised chromatin signature, was shown to be both temporally and spatially restricted in transgenic zebrafish embryos (Rada-Iglesias et al., 2011).

The binding of the transcriptional coactivator p300 was shown to correlate with the location of active enhancers in cell culture assays (Heintzman et al., 2007). The predictive power of this association was revealed in a study that assessed the genome wide binding of p300 in tissue isolated from the forebrain, midbrain and limb buds of E11.5 mouse embryos by ChIP-seq (Visel et al., 2009). The transcriptional activity of 86 p300 bound genomic regions was determined using a transgenic mouse assay, of which 87% were shown to function as enhancers (Visel et al., 2009). These data also revealed that the tissue specific binding of p300 was correlated with the activity of a given enhancer *in-vivo*, such that 78% of forebrain, 82% of midbrain and 80% of limb derived elements were active in their tissue of origin (Visel et al., 2009). Furthermore, approximately 90% of all ChIP-seq peaks were shown to overlap with regions that are conserved between the human, mouse, rat, chick and fugu genomes (Siepel et al., 2005 ; Visel et al., 2009).

Together, these investigations highlight the ability of high throughput genomics to identify the chromatin signatures of enhancers and their association with key

transcriptional co-activators genome wide. Both of these properties have been shown to be predictive of enhancer function *in-vivo* and can be used to predict the spatiotemporal activity of regulatory regions. The data produced from these techniques differs from that obtained from comparative genomics, as it is only applicable to the cell type in which the experiment is performed. Conversely, comparative genomics is sufficient to identify putative regulatory elements but cannot provide any indication of the tissue specificity or temporal activity of these regions *in-vivo*. The observation that 90% of all p300 ChIP-seq peaks reside in areas of measurable sequence conservation highlights the value of employing both comparative and post-genomic techniques to functionally annotate the genome, when possible.

1.2.4 – Enhancers underlie the regulatory logic of developmental processes

The process of embryogenesis is mediated by GRNs, which act to specify and maintain cell identity by coordinating the expression of transcription factors. Enhancers are central to this process as they provide a mechanistic link between the transcriptional state of a cell and its spatiotemporal gene expression profile. Consequently, one of the major goals of developmental biology is to understand the cis-regulatory logic that governs developmental processes. The DV patterning of the syncytial *Drosophila* embryo represents one of the best examples of how the modification of enhancer architecture is sufficient to establish the basis of differential gene expression during development. This system is based on the interpretation of a nuclear concentration gradient of the transcription factor Dorsal, which is induced ventrally within the embryo (Moussian and Roth, 2005). The Dorsal gradient regulates the expression of over 40 transcription factors in a concentration dependant manner to define the mesoderm, neurogenic ectoderm and the dorsal ectoderm of the early embryo. The expression profiles of Dorsal induced transcription factors are classified into a total of 6 spatial domains across the DV axis (Hong et al., 2008).

The regulatory logic that decodes the nuclear Dorsal gradient into distinct expression domains was determined in a series of bioinformatics and ChIP-chip analyses of the enhancers associated with DV markers (Markstein et al., 2002 ; Papatsenko and Levine., 2005 ; Zeitlinger et al., 2007). These studies revealed that genes expressed in the ventral presumptive mesoderm are induced by high levels of Dorsal, in concert with its early target gene Twist (Jiang et al., 1991). The enhancers associated with this class of genes are characterised by the presence of multiple low-

affinity Dorsal and Twist binding sites (Hong et al., 2008). In contrast, markers of the neurogenic ectoderm are induced by intermediate levels of Dorsal and low levels of Twist. All of the enhancers associated with this class of genes contain a conserved binding module consisting of a high affinity Dorsal site adjacent to an asymmetric Twist site. This structural arrangement of binding sites is crucial for the cooperative transcriptional activity of these regulators (Hong et al., 2008 ; Markstein et al., 2004). Dorsal ectoderm markers are induced by low levels of Dorsal in the absence of Twist, which is reflected in their CRMs by the presence of high affinity Dorsal sites (Hong et al., 2008). These 3 major classes of enhancers are elaborated upon by the addition of binding sites for spatially restricted transcriptional activators or repressors, which act in concert with Dorsal to generate a further 3 expression domains (Hong et al., 2008).

Studies investigating the cellular interpretation of the Dorsal gradient revealed many of the common properties that allow enhancers to coordinate spatiotemporal gene expression during development. These include the presence of sites with different affinities for transcription factor binding, a feature that is likely to be fundamental to the interpretation of morphogen gradients in the neural tube (Dessaud et al., 2008). The presence of binding modules that permit the stereotyped interaction of transcription factor complexes, such as those observed for Dorsal and Twist, provides a further mechanism by which a limited number of key regulators might coordinate differential gene activation within a pool of progenitors. Finally, the interaction of both positive and negative transcriptional regulators upon the same enhancer was shown to modify the expression domain of the associated gene. This property of enhancers is likely to be central to the molecular interpretation of the ventral neural tube GRN, which is built upon positive Shh input and locally expressed transcriptional repressors. Despite significant advances in the field, the cis-regulatory logic that underlies the activity of the neural tube GRN remains poorly described.

1.3 – Pax3 and Pax7 are key members of the neural tube GRN

Studies exploring the series of repressive interactions that form the basis of the neural tube GRN have focused upon the assignment and maintenance of ventral progenitor identity. These investigations have revealed the ability of this repressive network to interpret a gradient of Shh activity over time and confer robustness to neuronal patterning (Balaskas et al., 2012). By comparison, the regulatory logic that defines and maintains the identity of interneuron progenitors within the dorsal neural

tube is poorly understood. In the following section I will introduce the Pax family of transcription factors and the unique structural domains that define their protein products. I will outline the expression profiles of Pax3 and Pax7 (Pax3/7) within the dorsal neural tube before exploring their function and transcriptional regulation during the process of CNS patterning.

1.3.1 – The Pax gene family

The *paired* locus was identified in a seminal study of mutations affecting the segmental pattern of *Drosophila* larva (Nüsslein-Volhard and Weichaus, 1980). Mutation of this locus was shown to result in the deletion of alternating body segments within the embryo, the defining feature of the pair-rule class of developmental patterning genes (Nüsslein-Volhard and Weichaus, 1980). The *paired* gene was subsequently cloned (Kilchherr et al., 1986) and sequenced (Frigerio et al., 1986), correlating its expression profile with the dramatic effects of the mutation *in-vivo*. The protein produced from the *paired* locus was demonstrated to encode two DNA binding domains, an extended homeobox and a novel 128 amino acid paired box (Bopp et al., 1986). Studies performed in *Drosophila* determined that the *gooseberry* segment polarity gene also encoded a domain with homology to the paired box, the sequence of which was used to identify and clone the mouse *Pax* (paired box containing) *1* gene (Baumgartner et al., 1987 ; Deutsch et al., 1988). Biochemical studies subsequently demonstrated that the product of the paired box, the paired domain (PD), interacted with DNA independently of the paired-type HD, establishing the PD as the defining feature of a novel class of transcription factors (Treisman et al., 1991).

The majority of vertebrate genomes contain 9 Pax genes, all of which encode a N-terminal PD that identifies the family (Chi and Epstein, 2004). Analyses of the structures of the Pax6 and *paired* PD complexed with DNA revealed that the PD is composed of two helix-turn-helix subdomains, commonly referred to as PAI and RED (Jun and Desplan, 1996 ; Xu et al., 1995 ; Xu et al., 1996). The structure of the Pax6 PD demonstrated that it makes contacts with successive major grooves in the target sequence, such that the N-terminal PAI unit binds a 5’ half site whereas the C-terminal RED unit interacts with 3’ sequence (Xu et al., 1996). A series of biochemical studies defined the PD consensus binding site as (G/T)T(T/C)(C/A)(C/T)(G/C)(G/C) (Jun and Desplan, 1996), however individual Pax genes have been shown to have specific preferences for positions within this sequence (Czerny et al., 1993 ; Epstein et al 1994).

Furthermore, the oligonucleotides used to crystallize the *paired* and Pax6 PD were between 15-20 bp in length, suggesting that the binding site occupied by the PD may be significantly larger than the 7 bp represented by the consensus sequence (Xu et al., 1995 ; Xu et al., 1996). In addition to the PD, all Pax genes encode a C-terminal proline/serine/threonine rich weak transactivation domain that is thought to be partly responsible for the activation of target gene expression (Chalepakis et al., 1994 ; Glaser et al., 1994 ; Robichaud et al., 2004).

Despite sharing common features, vertebrate Pax genes are subdivided into four groups based on the composition of additional structural domains within their protein products (Fig. 1.4) (Balczarek et al., 1997). With the exception of the group IV genes, all Pax proteins contain a short octapeptide domain (Fig. 1.4). Deletion of this motif from Pax2 was shown to increase its transcriptional activity, suggesting that it might function as a transcriptional repressor (Lechner and Dressler, 1996). Sequence analysis of the octapeptide domain encoded by group II Pax genes revealed homology to the repression domains of the Engrailed and Goosecoid transcription factors, both of which are known to recruit members of the Groucho family of transcriptional repressors to DNA (Eberhard et al., 2000 ; Smith and Jaynes., 1996). Consistent with this sequence analysis, the octapeptide domain of group II genes has also been shown to directly interact with Groucho proteins, providing a molecular mechanism of Pax target gene repression (Eberhard et al., 2000 ; Patel et al., 2012).

Pax proteins are also classified according to the presence and type of HD they contain. Group I genes do not encode a HD and therefore rely upon the PD for sequence recognition (Chi and Epstein, 2002 ; Deutsch et al., 1988). Group III and IV encode an extended paired-type HD, whereas group II encode a truncated domain that contains only the first helix (Fig. 1.4). Both the full and partial HDs present within vertebrate Pax genes have been demonstrated to directly interact with the TATA-binding protein, linking Pax gene function to the direct recruitment of the basal transcription machinery (Cvekl et al., 1999 ; Eberhard et al., 1999). The DNA sequence preferentially occupied by the HD of Pax genes is defined as a palindrome of the canonical homeobox motif separated by 2-3 free positions in the form of TAAT(N_{2-3})ATTA, both by *in-vitro* selex (Wilson et al., 1993) and ChIP-seq (Soleimani et al., 2012). The presence of both a PD and a HD within the majority of Pax genes greatly increases the number of putative target sites within the genome, a feature that is further complicated by cooperative interactions between these DNA binding domains during sequence recognition (Jun and Desplan, 1996).

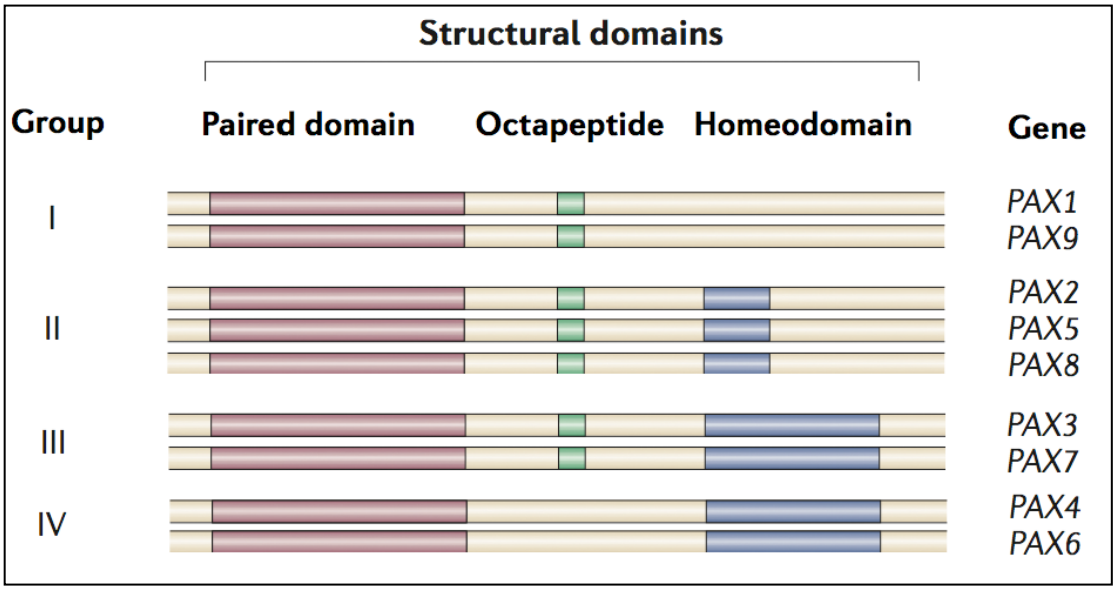


Fig. 1.4 – The vertebrate Pax gene family

The 9 vertebrate Pax genes are classified into 4 groups based on the composition of structural domains in their protein products. All genes encode a DNA-binding PD, which is the defining feature of the family. Group I, II and III proteins contain an octapeptide domain that is capable of directly recruiting Groucho family repressors to DNA. Group I do not encode a HD, whereas Group II encode a partial domain consisting of the first helix. Both group III and IV genes contain a full paired-type HD.

Figure adapted from Robson et al., 2006.

1.3.2 – Pax3 and Pax7 are expressed in the alar plate of the neural tube

Studies characterising the expression profiles of Pax genes during mouse embryogenesis revealed that both *Pax3* and *Pax7* are expressed within the progenitor cells of the dorsal spinal cord (Goulding et al., 1991; Jostes et al., 1990). Furthermore, these reports demonstrated that both genes share a sharp ventral boundary of expression coinciding with the sulcus limitans, a longitudinal widening of the lumen that divides the CNS into a dorsal alar and a ventral basal plate. This anatomical division of the embryonic CNS was first documented in pioneering histological studies performed by Wilhelm His in the late 19th century (His. 1888). As a result of these studies, His proposed that the sulcus limitans marked the lineage restriction of neuronal progenitors within the embryonic alar plate to sensory neuronal fates and those within the basal plate to motor associated fates (His. 1888). This theory is supported by subsequent investigations demonstrating the regionalised expression of cell adhesion molecules at the level of the sulcus limitans in the ventricular zone of the spinal cord. The dorsal limit of Cadherin-7 expression was shown to be localised at the sulcus limitans of the developing chick spinal cord between Hamburger and Hamilton stage (HH) 16-28 (Hamburger and Hamilton, 1951 ;Nakagawa and Takeichi, 1995 ; Nakagawa and Takeichi, 1998) and was also demonstrated to be coincident with the ventral boundary of Pax7 expression at HH stage 19 (Ju et al., 2004). Studies in *Xenopus* embryos demonstrated that *F-cadherin* is expressed at the sulcus limitans in the developing spinal cord and furthermore, that the expression of this cell adhesion molecule is initiated within the neural plate (Espeseth et al., 1995). These data begin to establish a molecular basis to the anatomical observations made by Wilhelm His and suggest that any lineage constraint imposed by the sulcus limitans may be enforced by restricted cell migration between the alar and basal plates.

Analysis of clonal dispersion during CNS development supports the hypothesis that the temporal expression of cell adhesion molecules in the neuroepithlium enforces restricted cell migration between the alar and basal plates. In one such study, donor cells expressing LacZ were transplanted into the presumptive alar or basal plate of a host *Xenopus* embryo prior to gastrulation and assessed following neurulation. These data demonstrated that clones made within either the presumptive alar or basal plates prior to neurulation readily disperse throughout these compartments but are restricted from crossing the sulcus limitans, consistent with the expression of *F-cadherin* in the *Xenopus* neural plate (Espeseth et al., 1995 ; Espeseth et al., 1998). A similar study of

clonal dispersion during zebrafish neurulation supported the hypothesis that progenitor cells are restricted to either the alar or basal plates prior to neurulation, consistent with the temporal expression profile of *cadherin-7* during zebrafish CNS development (Liu et al., 2007 ; Papan and Campos-Ortega, 1999).

In opposition to these conclusions, experiments performed in chick embryos suggested that the alar/basal division of the neuroepithelium is not established in the neural plate. Clones made in the open neural plate at HH stage 8 and assessed following neurulation at HH stages 19-29 were shown to readily disperse throughout the DV axis and cross the sulcus limitans (Erskine et al., 1998 ; Leber and Sanes, 1995). These studies also documented a progressive restriction of cell mixing in the chick ventricular zone following neurulation, such that cells labelled at HH stages 14-20 did not migrate between the alar and basal plates (Erskine et al., 1998 ; Leber and Sanes, 1990). The observed period in which cell mixing in the DV axis of the chick ventricular zone is restricted broadly correlates with the defined expression profile of Cadherin-7 at the sulcus limitans (Nakagawa and Takeichi, 1995 ; Nakagawa and Takeichi, 1998). Taken together, these reports suggest that the spatiotemporal expression of cell adhesion molecules within the neuroepithelium, rather that the anatomical formation of the sulcus limitans, might be the major determinant of progenitor cell dispersion during CNS development.

Both Pax3 and Pax7 are expressed in the dorsal spinal cord and exhibit a ventral boundary that coincides with the sulcus limitans, as previously described. Despite this shared boundary, the spatiotemporal dynamics of gene expression within the spinal cord is not conserved between these paralogous genes. *Pax3* transcript is first detected during neurulation, within the neural folds and along the closed dorsal neural tube of E8.5-9 mouse embryos (Goulding et al., 1991). By E10, expression is observed in all cells within the dorsal progenitor domains of spinal cord and the RP along the entire AP axis of the embryo. A sharp ventral boundary of gene expression at the level of the sulcus limitans was observed from E10 and was shown to be maintained at this DV position within the neural tube until E13, after which cells become committed to neuronal lineages and extinguish *Pax3* expression (Goulding et al., 1991).

Pax3 expression in chick embryos is initiated at a comparatively earlier stage of CNS development than reported in mice. Transcripts are first observed in the primitive streak at HH stage 7 and throughout the medio-lateral axis of the neural plate caudal to the node between HH8-9 (Bang et al., 1997 ; Goulding et al., 1993). Expression is extinguished in medial regions of the neural plate by HH stage 10 and is restricted to the

tips of the neural folds during neurulation (Goulding et al., 1993 ; Liem et al., 1995). Serial sectioning of embryos at HH stage 10 revealed that that Pax3 expression expands ventrally to reach its final limit at a point midway between the roof and floor plates (Goulding et al., 1993).

Pax7 expression is first detected in the neural tube following neurulation at E9 in mice (Jostes et al., 1990). Transcription is observed in the dP2-dP6 progenitor domain and shares a ventral limit with the sulcus limitans, however it is excluded from the dP1 domain and the RP (Jostes et al., 1990). This expression profile in the spinal cord is evident between E9-E13 in the dorsal spinal cord, after which progenitor cells undergo differentiation and *Pax7* transcript is no longer detected. These data indicate that *Pax7* is not a direct molecular correlate of the alar plate in mouse embryos, despite sharing a ventral boundary of expression with both the sulcus limitans and Pax3. Expression studies in chick revealed a comparable temporal profile of Pax7 transcription to that observed during in mouse embryogenesis. Accordingly, Pax7 is first expressed in the dorsal neural tube following neurulation at HH stage 10 (Kawakami et al., 1997). In contrast to the observed expression profile in the spinal cord of mice, Pax7 was shown to be expressed throughout the entire alar plate of chick embryos (Kawakami et al., 1997).

1.3.3 – The function of Pax3/7 in the neural tube

Despite their broad expression domain within the neural progenitors cells of the dorsal spinal cord, the functions of Pax3/7 during CNS patterning and neurogenesis are poorly defined. Analysis of the naturally occurring *splotch* allele in mice demonstrated that this mutation was homozygous lethal at approximately E14, with embryos exhibiting severe neural tube defects including a failure to complete neurulation both anteriorly and posteriorly (Auerbach, 1954). Subsequent studies cloned the *splotch* mutation to the *Pax3* locus, establishing this gene as a putative regulator of neural tube closure (Epstein et al., 1991 ; Epstein et al., 1993). Loss of *Pax3* expression during neurulation was shown to result in the activation of apoptosis within the neural folds, suggesting that the primary function of *Pax3* in this process is to facilitate cell survival (Phelan et al., 1997). Consistent with this hypothesis, knockdown of the *p53* tumor suppressor gene, either genetically or pharmacologically, was sufficient to rescue the neural tube defects observed in *splotch* homozygotes to wild type ratios (Pani et al., 2002).

Mice carrying a targeted insertion of *LacZ* within the *Pax7* locus, resulting in the loss of transcript in homozygous offspring, do not exhibit any defects in neurulation (Mansouri et al., 1996). The lack of a neural tube defect in these mice was reasoned to be due to the expression of *Pax3* at an earlier stage within the neuroepithelium, which may compensate for the loss of *Pax7* following neural tube closure (Goulding et al., 1991 ; Jostes et al., 1990 ; Mansouri et al., 1996). This hypothesis was tested in experiments that assessed the functional redundancy between Pax3 and Pax7 by replacing the *Pax3* coding sequence with *Pax7* cDNA in transgenic mice (Relaix et al., 2004). This investigation clearly demonstrated that Pax3/7 perform redundant functions during neurulation and furthermore, that *Pax7* cDNA expressed under the control of *Pax3* regulatory elements can rescue neural tube closure in a *splotch* mutant background (Relaix et al., 2004).

The observed functional redundancy of Pax3/7 within the spinal cord has limited the opportunity to study the requirement of these genes during CNS patterning and neurogenesis, as these investigations require the generation of double mutant mice. These constraints have resulted in the publication of a single report documenting the distribution of interneuron subtypes in the spinal cord of *Pax3;Pax7* mutant embryos (Mansouri and Gruss, 1998). These embryos exhibit a severe neural tube closure defect and die before E11, however the authors reported that the transcription of the RP derived signaling molecules Wnt1, Wnt3 and Wnt3a remained unchanged (Mansouri and Gruss, 1998). The expression domain of the V1 interneuron marker En1 was shown to be expanded dorsally, suggesting that loss of both Pax3/7 results in a de-repression of ventral neuronal fates (Mansouri and Gruss, 1998). However, this shift was not recapitulated by the V0 marker Evx1, which was limited to its endogenous domain (Mansouri and Gruss, 1998). The generation of a novel *splotch* allele by *LacZ* insertion to the *Pax3* locus enabled the authors to demonstrate that *Pax3* is expressed in a population of commissural interneurons within the dorsal spinal cord (Mansouri and Gruss, 1998). A similar expression profile of *LacZ* labelling was observed in mice carrying the previously described targeted *Pax7* mutant allele (Mansouri et al., 1996 ; Mansouri and Gruss, 1998). LacZ expression within these commissural neurons was lost in *Pax3/Pax7* mutant embryos, resulting in the conclusion that both *Pax3* and *Pax7* are required for the specification of this population of dorsal interneurons (Mansouri and Gruss, 1998). These data provided evidence that Pax3/7 are required for the repression of ventral neuronal identities and the specification of dorsal subtypes,

however the patterning mechanisms that underlie these phenotypes were not extensively investigated during this study.

A more extensive analysis of the role of *Pax3* in the acquisition of progenitor identity was performed in transgenic mice expressing *Pax3* cDNA under the control of the *Hoxb-4* enhancer A region (Tremblay et al., 1996 ; Whiting et al., 1991). This experimental manipulation resulted in the expression of *Pax3* throughout the DV axis of the spinal cord in regions posterior to the forelimb (Tremblay et al., 1996). Transgenic embryos exhibited an expansion of *Pax7* expression into the ventral half of the spinal cord at E12.5, however the distribution of transcript did not extend to the RP dorsally or beyond the pMN domain ventrally (Tremblay et al., 1996). These data suggest that the repression of *Pax7* transcription within these domains cannot be overcome by *Pax3* overexpression. In contrast, *Pax6* transcripts were observed at the ventral midline of the neural tube (Tremblay et al., 1996). Ectopic *Pax3* transcription was shown to completely extinguish the expression of *Shh* and *HNF3β* (Foxa2), indicating that either *Pax3* or *Pax6* possess the ability to repress the key determinants of floor plate identity (Tremblay et al., 1996). Finally, the authors demonstrated that transgenic mice exhibited a 50% reduction in the number of motor neurons at E13.5, consistent with the lack of molecular markers of the floor plate and the dorsalisation of the neuroepithelium (Tremblay et al., 1996). These data demonstrate that Pax3 expression imparts a dorsal identity upon neuronal progenitor cells but do not provide any mechanistic insights into the function of this gene within neural progenitors.

Analysis of neural crest explant cultures prepared from *splotch* mutant mice at E10 demonstrated that these cells undergo increased neurogenesis when compared to wild type controls (Nakazaki et al., 2008). The converse result was obtained in explants taken at E12.5, suggesting that the premature neurogenesis in these cultures does not persist due to progenitor depletion (Nakazaki et al., 2008). These data are in agreement with a previous report documenting a significant reduction in neurogenesis within *splotch* explant cultures assessed at E12 (Koblar et al., 1999). *In-situ* hybridisation experiments revealed that *splotch* mutant embryos lacked expression of *Hes1*, a key effector of the notch pathway that is involved in the maintenance of cells in a progenitor state (Nakazaki et al., 2008 ; Kageyama and Ohtsuka, 1999). The authors also demonstrated that expression of *Ngn2*, a bHLH transcription factor involved in the acquisition of neuronal fates, was absent in the spinal cord of *splotch* mutants (Nakazaki et al., 2008 ; Bertrand et al., 2002). Furthermore, the regulation of both *Hes1* and *Ngn2* by *Pax3* was shown to be direct in both ChIP and electrophoretic mobility shift assays

(EMSAs) (Nakazaki et al., 2008) These data led the authors to hypothesise that *Pax3* functioned within the dorsal neural tube to coordinate the balance of proliferation and differentiation within neural progenitors.

A subsequent study proposed that the acetylation state of *Pax3* underlies the dual function of this transcription factor during CNS development (Ichi et al., 2010). At E9.5, deacetylated *Pax3* was shown to activate *Hes1* transcription and repress *Ngn2*, promoting the maintenance of cells in a progenitor state (Ichi et al., 2010). Conversely, acetylated *Pax3* was observed to induce *Ngn2* expression at E12.5 and repress transcription from the *Hes1* promoter, facilitating cell cycle exit and neurogenesis (Ichi et al., 2010). The temporal switch in the acetylation state of *Pax3* was shown to be regulated by the activity of the histone deacetylase SIRT1, which was associated with both the *Hes1* and *Ngn2* promoters at E9.5 but absent from both at E12.5 (Ichi et al., 2010).

Taken together, these data provide the first mechanistic insights into the function of Pax3 expression in spinal cord progenitors. Furthermore, they suggest that this transcription factor may primarily act to temporally regulate the balance of proliferation and differentiation within neuronal progenitors, consistent with its well described roles in neural crest and muscle development (Lang et al., 2005 ; Buckingham and Relaix, 2007).

1.3.4 – Regulation of Pax3/7 expression by morphogens

The establishment of the Pax3/7 expression domain is one of the earliest patterning events to occur during neural tube development and molecularly divides the tissue into dorsal and ventral domains. This is a key event in the establishment of ventral progenitor identities within the spinal cord and is therefore commonly used to assess the state of the neural tube GRN in response to manipulations of morphogen signaling pathways. In addition to providing mechanistic insights into the role of morphogens in neural tube patterning, these experiments are also informative when interpreted in the context of Pax3/7 regulation.

The Shh pathway is the major determinant of progenitor identity within the ventral neural tube and consequently represents a good candidate to play a role in the establishment of the Pax3/7 expression domain. The first study to implicate this signaling pathway in the regulation of Pax3/7 expression removed the notochord from a defined region of the chick embryo following neurulation and assessed the effect of this

manipulation on the expression of Pax3 and Pax6 (Goulding et al., 1993). These data revealed that notochord removal results in the failure of FP induction and a prominent ventral expansion of both Pax3 and Pax6, indicating the neural tube was dorsalised in the absence of Shh (Goulding et al., 1993). In the converse manipulation, the authors grafted explants of either notochord or FP tissue to one side of the neural plate and demonstrated that that the transcription of both Pax3 and Pax6 was repressed (Goulding et al., 1993). Direct evidence supporting the requirement of the Shh pathway in the ventralisation of the neural tube was subsequently provided in a study that determined the response of neural plate explants to pathway manipulation *ex-vivo* (Ericson et al., 1996). These data revealed that 95% of cells within an explant of neural plate tissue taken at HH stage 10 express Pax7 and that this expression is extinguished if either notochord explants, or exogenous Shh is added to the culture (Ericson et al., 1996). Furthermore, electroporation experiments in which activated forms of the Gli transcription factors are expressed throughout the neural tube revealed the cell autonomous repression of Pax7 expression within its endogenous domain (Stamataki et al., 2005). Together, these investigations established the basis of a model in which Shh signaling is required to repress Pax3/7 during the early stages of neural tube patterning to ventralise the tissue.

Mouse models in which Shh itself or key components of the signal transduction pathway are mutated provide support for this model of Pax3/7 regulation in the neural tube. *Shh*$^{-/-}$ embryos were shown to exhibit an expansion of both *Pax3* and *Pax7* expression throughout the entire DV axis of the neuroepithelium, highlighting the key role of the ligand in ventral patterning (Litingtung and Chiang, 2000 ; Wijgerde et al., 2002). Clonal loss of the pathway transducer *Smo* was shown to result in the cell autonomous induction of Pax7 expression within the ventral domains of the spinal cord (Wijgerde et al., 2002). Futhermore, electroporation of a dominant inhibitory form of the Gli3 transcription factor in the chick neural tube revealed a cell autonomous induction of Pax7 expression in the ventral spinal cord (Meyer and Roelink, 2003 ; Persson et al., 2002).

The hypothesis that Shh signaling acts to establish the expression domain of Pax3/7 is complicated by the results of studies in which both the positive and negative regulators of the pathway are inactivated. *Shh*$^{-/-}$*;Gli3*$^{-/-}$ mutant mice, in which both the ligand and the negative regulator of the pathway are mutated, exhibit a ventral boundary of *Pax3* and *Pax7* expression at the level of the sulcus limtans (Litingtung and Chiang, 2000 ; Persson et al., 2002). In agreement with this result, loss of both *Smo* and *Gli3*

was shown to restore the ventral boundary of *Pax7* expression to its endogenous position (Wijgerde et al., 2002). These data indicate that the Pax3/7 domain is established at the correct DV position within the neural tube in the absence of all forms of hedgehog signaling. Further evidence in support of this hypothesis was provided by analysis of spinal cord patterning in mutants affecting the structure and function of the primary cilium, which has been shown to be required for all Shh signaling in vertebrates (Goetz and Anderson, 2010). Mice carrying mutations in ift88 (*polaris*), ift172 (*wimple*), or C2cd3 (*Hearty*) exhibit a complete loss of Shh signaling in the neural tube but express *Pax7* within its endogenous domain (Hoover et al., 2008 ; Huangfu and Anderson, 2003).

The role of the Wnt signaling pathway in the positive regulation of Pax3 expression was established in experiments performed in chick and *Xenopus* embryos (Bang et al., 1997 ; Bang et al., 1999). The authors demonstrated that the posterior non-axial mesoderm of HH stage 8 chick embryos was sufficient to induce Pax3 expression in neuralised *Xenopus* animal caps (Bang et al., 1997). Subsequent experiments determined that the ability of theis tissue to induce Pax3 expression is dependent on the activity of Wnt8 (Bang et al., 1999). The expression profiles of Pax3 and Wnt8 were shown to overlap in the *Xenopus* presumptive neural plate and furthermore, over expression of Wnt8 resulted in the expansion of Pax3 transcription (Bang et al., 1999). The converse experiment revealed that knockdown of Wnt8 in the neural plate extinguishes Pax3 expression (Bang et al., 1999). These data suggest that the Wnt signaling pathway is both necessary and sufficient for Pax3 expression in the lateral neural plate.

Subsequent investigations have built upon this hypothesis by demonstrating that ectopic expression of Wnt1 and Wnt3a in the chick spinal cord is sufficient to expand the Pax7 domain at the expense of ventral progenitor markers on the electroporated side (Alvarez-Medina et al., 2008). A similar expansion of the Pax7 domain is observed in electroporation experiments employing a constitutively active form of the Wnt pathway effector β-catenin (Alvarez-Medina et al., 2008 ; Zechner et al., 2007). In contrast, mouse embryos in which the transcriptional activity of β-catenin is specifically inactivated exhibit a strong reduction of Pax3 expression in the dorsal spinal cord at E12.5 (Valenta et al., 2011). The requirement of the Wnt signaling pathway for Pax3/7 expression is further highlighted in chick studies employing a dominant negative form of the Wnt pathway effector Tcf3, which revealed a cell autonomous repression of both

Pax3 and Pax7 within their endogenous domains (Alvarez-Medina et al., 2008 ; Lee and Deneen, 2012).

These data are consistent with the hypothesis that the Wnt signaling pathway is both necessary and sufficient for Pax3/7 expression in the neural tube, however there is also evidence to suggest that this interaction may be indirect. Overexpression of Wnt1 and Wnt3a in the chick neural tube was shown to result in the induction of Gli3 transcription and when this interaction is blocked, the Wnt pathway can no longer induce Pax7 expression (Alvarez medina et al., 2008). Furthermore, either electroporation of dominant negative Tcf3 in chick or *Wnt1*$^{-/-}$*;Wnt3a*$^{-/-}$ mutant mice was shown to result in the loss of Gli3 expression in the dorsal spinal cord (Alvarez medina et al., 2008). Together, these data suggest that the interaction of the Shh and Wnt signaling pathways within the neural tube might play a role in the regulation of Pax3/7 expression.

The BMP signaling pathway is also thought to positively regulate the expression of Pax3/7 in CNS progenitors, however this has not been as extensively studied. The role of this pathway in Pax3/7 regulation was established in a series of *ex-vivo* experiments that demonstrated the ability of the epidermal ectoderm to induce Pax3 transcription in the underlying neural plate (Liem et al., 1995). The identities of the inducing factors were subsequently defined as BMP4 and BMP7 (Liem et al., 1995). Experiments investigating the role of BMP signaling in the specification of progenitor identity in the chick spinal cord revealed that retroviral mediated expression of an activated form of BMP receptor 1b throughout the neural tube results in an a slight ventral expansion of Pax7 on the electroporated side (Timmer et al., 2002). Comparatively greater expression levels of BMP receptor 1b, driven by the cytomegalovirus promoter produce a greater expansion of Pax7 expression, however these results are difficult to interpret as apoptosis is induced in this experiment (Timmer et al., 2002). In contrast to these conclusions, knockdown of BMP7 in the chick spinal cord has been shown to have no effect on Pax7 expression (Le Dréau et al., 2012). Similarly, RNAi mediated knockdown of Smad1 and Smad5 in chick did not reduce the Pax7 expression domain (Le Dréau et al., 2012). Finally, the boundary of Pax7 expression in *BMP7*$^{-/-}$ mice was shown to be identical to wild type (Le Dréau et al., 2012). Taken together, these reports suggest that the BMP signaling pathway is sufficient but not necessary for the expression of Pax3/7.

1.3.5 – The molecular basis of Pax3/7 regulation in the neural tube

Several studies have attempted to define the regulatory mechanisms that act to initiate Pax3 transcription in the neural plate and restrict its activity to the dorsal half of the spinal cord, all of which are focused upon a 1.6 kilobase (kb) region located upstream of the Pax3 transcription start site in the mouse genome. The first report to suggest that this region possessed a regulatory potential was based on the differentiation of P19 embryonal carcinoma (EC) cells into either Pax3 expressing neuroectoderm or mesoderm, achieved by culture in the presence of RA or DMSO, respectively (Natoli et al., 1997 ; Pruitt. 1994). A 14 kb fragment of 5' Pax3 sequence, fused to a chloramphenicol acetyltransferase reporter gene, was shown to activate transcription in either RA or DMSO treated cells (Natoli et al., 1997). Serial deletion of this fragment revealed that 1.6 kb of sequence located proximal to the Pax3 promoter retained the ability to induce reporter expression in cells differentiated into either neuroectoderm or mesoderm (Natoli et al., 1997).

Transgenic mice expressing *LacZ* under the control of either the 1.6 kb or 14 kb fragments of upstream sequence demonstrated that these regulatory regions were transcriptionally active at E8.5 in the neural plate and recapitulated the dorsal restriction of Pax3 expression within the spinal cord (Natoli et al., 1997). The authors noted that the expression mediated by the 1.6 kb construct was observed throughout the mesoderm of the tail bud, in contrast to *Pax3* transcription, which is restricted to the neural plate (Natoli et al., 1997). This observation led the authors to conclude that the 1.6 kb region located proximal to the Pax3 transcriptional start site was sufficient to recapitulate gene expression in the dorsal spinal cord, however additional elements located within the 14 kb fragment were required to repress Pax3 expression from the tail bud mesoderm (Natoli et al., 1997).

A series of EMSA experiments, performed using a series of overlapping fragments within the 1.6 kb CRM, identified 4 sites in the element occupied by nuclear extract prepared from P19 EC cells differentiated to neuroectoderm (Pruitt et al., 2004). Subsequent analysis of each individual site by EMSA demonstrated that 2 sites, termed sites I and II, were bound by the POU class homeobox genes Brn1 and Brn2 (Pruitt et al., 2004). Sites A and B were shown to interact with the Hox cofactors Pbx and Meis1/2 (Pruitt et al., 2004). Mutation of either site I or site II in the context of the 1.6 kb Pax3 CRM was shown to reduce the level of reporter transcription in transgenic mice, however the domain of expression within the dorsal neural tube was retained

(Pruitt et al., 2004). Similarly, mutation of site A was also demonstrated to have no effect upon transgene expression *in-vivo*. LacZ expression was observed to be downregulated in the anterior regions of the neural tube of site B mutants at E9.5 and completely extinguished in the tissue by E11.5 (Pruitt et al., 2004). These results led to the conclusion that the Pbx/Meis motif represented by site II modulates the expression of Pax3 in the AP axis of the CNS (Pruitt et al., 2004).

A subsequent investigation assessed the conservation of the 1.6 kb Pax3 CRM between the human and mouse genomes to identify two regions, named neural crest enhancers (NCE) 1 and 2 (Millewski et al., 2004). Removal of either of these regions from the 1.6 kb fragment was shown to result in a complete loss of transgene expression in mice at E10.5, suggesting that both are independently required for CRM activity (Millewski et al., 2004). A 454 bp construct containing only NCE1 and 2 was shown to recapitulate Pax3 expression in the somites, however activity within the spinal cord was dorsally restricted in comparison to endogenous transcript and did not exhibit a sharp ventral boundary (Millewski et al., 2004). The authors used the mouse sequence of NCE2 as bait in a yeast one-hybrid screen to identify Tead2, a transcriptional effector the Hippo signaling pathway whose expression profile overlaps with Pax3 in the dorsal spinal cord (Millewski et al., 2004). Tead2 was shown to bind NCE2 via a consensus Tead motif in EMSAs and induce transcription from NCE2 in luciferase assays (Millewski et al., 2004). Expression of a dominant negative form of Tead2 under the control of the *Wnt1* promoter revealed a slight reduction in Pax3 expression in the neural tube at E10.5, however these data are difficult to interpret as the morphology of the neuroepithlium was severely affected in this experiment (Millewski et al., 2004). A subsequent report assessing the requirement of Tead2 during neurulation clearly demonstrated that Pax3 transcription is not affected in *Tead2*$^{-/-}$ mice (Kaneko et al., 2007).

A recent study assessed the requirement of both NCE1 and NCE2 for *Pax3* expression in mice by replacing the region containing these elements with a neomycin resistance gene by Cre mediated recombination (Degenhart et al., 2010). Homozygous transgenic embryos exhibited severe neural tube defects, a lack of limb musculature and a dramatic reduction in *Pax3* expression within the dorsal spinal cord, suggesting that removal of NCE1 and NCE2 extinguished Pax3 expression *in-vivo* (Degenhardt et al., 2010). However, subsequent experiments demonstrated that targeting the neomycin resistance gene to the locus resulted in the creation of a null *Pax3* allele and furthermore, that removal of this cassette by homologous recombination completely

restored *Pax3* expression throughout the embryo (Degenhardt et al., 2010). These data clearly demonstrated that both NCE1 and NCE2 are not required for the induction of Pax3 transcription in the neuroepithelium or its restriction to the dorsal half of the spinal cord (Degenhardt et al., 2010). This led the authors to propose that the Pax3 locus contains multiple, functionally redundant CRMs that act in concert to define the Pax3 expression domain within the dorsal spinal cord (Degenhardt et al., 2010).

A comparative genomics investigation of the Pax3 locus, based on the conservation of non-coding sequence between the mouse and chicken genomes, revealed the presence of 8 evolutionary conserved regions (ECRs) with at least 50% identity over at least 200 bp (Degenhardt et al., 2010). A 967 bp element located within the 4th intron of the locus, named Pax3-ECR2, was shown to be sufficient to induce transcription within the neural plate of mice assessed at E8.5, consistent with the temporal induction of *Pax3* expression in the CNS (Degenhardt et al., 2010 ; Goulding et al., 1991). The authors also demonstrated that transgene expression mediated by ECR2 was localised to the dorsal spinal cord of mice between E9.5 – E12.5 in a profile consistent with that observed for *Pax3* (Degenhardt et al., 2010). ECR2 was shown to be transcriptionally active in zebrafish embryos at 24 hours post fertilisation (hpf) with an expression profile that overlapped with that of Sox10, a marker of the neural crest lineage (Degenhardt et al., 2010). Furthermore, the activity of ECR2 was shown to be dependent upon the integrity of 5 Tcf/Lef binding sites, which when mutated in the context of a 403 bp minigene resulted in a complete loss of transgene activation in zebrafish (Degenhardt et al., 2010).

These data were the first to provide a mechanistic link between the effects of Wnt pathway manipulation on Pax3 expression and the molecular basis of gene regulation. This study also revealed that, at least within the mouse genome, the expression profile of Pax3 in the CNS is governed by the activity of several functionally redundant cis-regulatory modules. However, the regulatory logic that defines the ventral boundary of the Pax3 expression domain was not explored during this investigation

1.4 – Aims

The expression profile of Pax3 within the embryonic CNS and the results of studies describing its function suggest that this transcription factor might play a critical role in the development of dorsal interneuron progenitors within the alar plate. Despite being one of the first patterning genes expressed by neural progenitors, the molecular basis of Pax3 transcriptional initiation within the lateral neural plate is poorly understood. Similarly, the regulatory mechanisms that produce a sharp boundary of expression at the level of the sulcus limitans cannot be described by a cross-repressive interaction, as no ventrally expressed genes share a boundary with Pax3 in the neural tube. This study aims to address these issues by determining the position of Pax3 within the neural tube GRN.

This investigation will begin by testing the hypothesis proposed by Wilhelm His, which states that alar plate progenitors are fated to give rise to sensory neurons in the adult tissue (His. 1888). This will be be achieved by employing Cre/LoxP technology in transgenic mice to trace the lineage of Pax3 expressing progenitors from the early stages of neural tube patterning to the establishment of functional sensory-motor connections at birth. These data will also be utilised to study the mechanisms through which a sharp boundary of Pax3 expression is produced in the spinal cord by comparing the distribution of Pax3 protein to cells derived from this lineage.

The molecular basis of Pax3 expression in the spinal cord will be examined in a comparative genomics study; which will identify CNEs associated with the locus. Transgenesis experiments, performed in zebrafish and chick embryos, will be employed to determine the regulatory potential of each CNE *in-vivo*. These studies will aim to reduce the complex profile of Pax3 expression to the enhancer activity of discrete CRMs. CNEs which exhibit transcriptional activity within spinal cord progenitors will be selected for functional dissection, guided by the *in-silico* annotation of regulatory motifs and TFBS within these regions. Finally, the interactions of regulatory sequences with putative transcription factors will be investigated *in-vitro* in order to determine the regulatory logic that governs Pax3 expression in the neural tube.

Chapter 2 – Materials and Methods

2.1 – General molecular biology techniques

2.1.1 – Transformation of chemically competent bacteria

1-5 µl of plasmid DNA was added to a 50 µl aliquot of lab stock chemically competent DH5-α bacteria and incubated on ice for 15 mins. The mixture was heat shocked at 42 ºC for 45 seconds and returned to ice for 2 mins. 1 ml of Luria-Bertani (LB) medium was added to the cells, which were then incubated at 37 ºC for 1 hour on a shaker. The transformed cells were then amplified in a 250 ml culture or spread onto LB agar plates for the identification of single colonies under the appropriate antibiotic selection.

2.1.2 – Plasmid DNA preparation

Large-scale DNA preparations of 250ml bacteria cultures were performed using the Qiafilter plasmid Maxi kit (Qiagen) according to manufacturers instructions. Small-scale preparations of 6ml bacterial cultures were performed using the Qiaprep spin miniprep kit (Qiagen). The concentration of all DNA preparations was determined using a Nanodrop ® spectrophotomer (LabTech) and supporting software before storage at 4 ºC.

2.1.3 – BAC DNA preparation

A 250 ml culture inoculated with a sample from a BAC glycerol stock and the appropriate antibiotic was incubated overnight at 30 ºC. DNA preparations were performed using a NucleoBond BAC 100 kit (Machery Nagel) according to the manufacturers instructions. The integrity of DNA was assayed by gel electrophoresis before the concentration was determined by spectrophotometry. BAC preparations were stored at 4 ºC and handled carefully to minimize the risk of DNA shearing.

Luria-Bertani media (1L) -10 g Tryptone, 5 g Yeast Extract, 10 g NaCl, pH 7.0
Ampicillin – 100 µg/ml
Kanamycin – 50 µg/ml

2.2 – Defining putative regulatory elements

2.2.1 – Comparative genomic alignments

Human genomic sequence (Hg18) containing the desired locus and all surrounding intergenic sequence was locally aligned to the equivalent regions in the mouse (Mm9), zebrafish (Zv5) and fugu (Fr2) genomes. Chromosome location and feature annotation was exported from the UCSC genome browser (http://genome.ucsc.edu/) and uploaded to the Mulan alignment tool (http://mulan.dcode.org/) with the appropriate repeat masking parameters selected. The computed phylogenetic tree was accepted as it accurately reflected the evolutionary relationships between the input genomes. Sequences were forwarded to the multiple threaded blockset alignment algorithm of the Mulan tool and the output was displayed as a graphical conservation profile. The preliminary alignment was refined to display CNEs that were present in all of the aligned sequences with at least 65% homology over at least 40 bp. The entire alignment was then restructured in order to map the location of CNEs to the zebrafish genome. 100 bp of genomic sequence was added to each end of these CNEs before the sequence was extracted.

2.2.2 – Cloning CNEs

Conserved regions were amplified from BAC preparations using the Expand High Fidelity Polymerase chain reaction (PCR) System (Roche). The optimal Oligonucleotides used for PCR amplification were selected by forwarding the sequence of CNEs to the Primer3 tool (http://biotools.umassmed.edu/bioapps/primer3_www.cgi). Primer pairs were designed to initiate amplification in the 100 bp of non-conserved genomic sequence added to each end of the CNE of interest. Each primer pair was modified to introduce a HindIII site at the 5' end of the amplicon and an SbfI site at the 3' end to facilitate ligation proximal to the promoter in the MiniTol2 TKProm Gal4-UAS Citrine (MiniTol2-Citrine) expression vector. The primers used to clone each CNE from the Pax3 and Pax7 loci are listed in Table 2.1.

Successful amplification of CNEs was confirmed by gel electrophoresis before the correct sized reaction product was excised with a scalpel on a long wave light box (365 nm). DNA was extracted from the gel using a Qiaquick gel extraction kit (Qiagen) according to manufacturers instructions and eluted in a volume of 20µl. Both the vector and insert were digested with HindIII and SbfI for 2 hours at 37 ºC before the vector was dephosphorylated with calf intestinal phosphatase for a further 30 mins to reduce the probability of re-ligation. The amount of insert DNA required in each ligation was calculated based on a 10:1 ratio of insert to vector in the reaction. A 20 µl reaction containing 50 ng of dephosphorylated vector, 1x T4 ligase buffer, 5U of T4 DNA ligase (Roche) and the required amount of insert was incubated for 1 hour at room temperature. 5 µl of the ligation mixture was transformed and spread onto LB agar plates for the identification of single colonies. Successful cloning was confirmed by sequencing the DNA preparations from at least 10 colonies using the MiniTol2 3' reverse primer listed in Table 2.1.

2.2.3 – Identifying CNE homology regions

The coverage of comparative genomic alignments was expanded by the identification of homology regions in additional vertebrate genomes. The Human sequence representing the minimal conserved region was used as the query sequence for all Basic Local Alignment Search Tool (BLAST) searches using the nucleotide BLAST algorithm of the Ensembl genome browser (http://www.ensembl.org/). The search was optimized for the identification of matches in distant genomes using the predefined settings accessed in the search sensitivity tab of the input page. The genomic location of BLAST hits was verified before the sequence was extracted and stored.

2.2.4 – Defining putative binding sites

The variance within CNE sequences present in available vertebrate genomes was analysed using the DNA function of the ClustalW2 multiple sequence alignment tool (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Expanding the width of the alignment facilitated the identification of blocks of single nucleotides that are highly conserved across numerous vertebrate genomes, these regions were hypothesized to represent candidates for functional TFBS.

Conserved motifs within putative regulatory regions were identified by comparative genomic analysis of homologous sequences using the MEME suite (http://meme.nbcr.net/meme/cgi-bin/meme.cgi). Statistically overrepresented motifs with a minimum width of 6 bp and a maximum of 15 bp that were present at least once in each of the query sequences were identified. Motifs were displayed as a consensus sequence and a position specific probability matrix (PSPM), calculated from the occurrence of each nucleotide within the motif across all input sequences. The PSPM representing each identified motif was compared to annotated TFBS in the Jaspar (http://jaspar.cgb.ki.se/) and uniPROBE (http://thebrain.bwh.harvard.edu/uniprobe/) databases using the TOMTOM search tool (http://meme.nbcr.net/meme4/cgi-bin/tomtom.cgi). TFBS databases were queried for similarity to discovered motifs using the Pearson correlation coefficient function and a significance threshold of E<10. The results were visualized as a graphical alignment of the predicted TFBS to the query motif and ranked according to probability.

2.2.5 – CNE modification by mutagenic PCR

15 bp evolutionary conserved motifs were deleted from CNEs using the Quickchange *II XL* site-directed mutagenesis kit (Stratagene) and the oligonucleotides listed in Table 2.3. The PCR strategy utilized complimentary primers that spanned the region of interest but omitted the central 15 bp, resulting in the production of a template in which the motif was not present. The manufacturers suggested cycling parameters were modified in order to increase the efficiency of the reaction, the optimized programme is outlined in Table 2.2. 10U of DpnI was used to digest the methylated template DNA before successful mutagenesis was visualized by gel electrophoresis. 5 µl of the reaction was transformed and spread onto LB agar plates and at least 10 colonies were sequenced for each motif deletion performed.

Targeted disruption of conserved TFBS within functional motifs was also performed using the Quickchange II XL site-directed mutagenesis kit (Stratagene). Nucleotides within TFBS identified by TOMTOM searches were mutated according to published data where available. The PSPM defining the TFBS of interest was used to design mutations in cases where published data was not accessible. The primers listed in Table 2.4 were used to introduce defined changes in CNEs using the cycling parameters outlined in Table 2.2. DNA preparations from at least 10 individual colonies were sequenced to confirm successful mutagenesis.

2.3 – Identifying DNA - protein interactions

2.3.1 – *In-vitro* transcription and translation

The full-length cDNA clones of the transcription factors listed in Table 2.5 were used as templates for *in-vitro* protein synthesis. Proteins were produced using the TnT coupled rabbit reticulocyte lysate system (Promega) in a 50 µl reaction containing 27.5 µl of rabbit reticulocyte lysate and 1µg of DNA template, according to manufacturers instructions. The reactions were snap frozen in liquid nitrogen upon completion and stored at -80 ºC until required.

2.3.2 – Radiolabelling of double stranded oligonucleotides

The single stranded oligonucleotides listed in Table 2.6 were resuspended in dH2O to a final concentration of 100 mM. Complementary Oligonucleotide pairs were mixed in a 1:1 ratio with 10x annealing buffer in a total volume of 100 µl. The mixture was heated to 95 ºC for 5 mins in a hot block before being allowed to return to RT over a period of 1 hour. Successful annealing was confirmed by electrophoresis on a 4% agarose gel before stocks were stored at -20 ºC.

The 5' end of annealed oligonucleotides was radiolabelled by the addition phosphorus 32 (^{32}P) using T4 polynucleotide kinase. A 20 µl reaction containing 500 ng of oligonucleotide, 1x T4 polynucleotide buffer (Promega), 10U of T4 polynucleotide kinase (Promega) and 2 µl of ^{32}P-γATP (3000 ci/mmol at 10 mCi/ml) (Hartman Analytic) was incubated at 37 ºC for 30 mins. 180 µl of TE was added to the reaction before purification on an illustra microspin G-25 column (GE Healthcare) according the manufacturers instructions. Radiolabelled oligonucleotides were stored at 4 ºC and used within one half-life.

2.3.3 – Chick nuclear extractions

36 fertilised chick eggs were incubated at 37 ºC and allowed to develop until embryonic day 3. Embryos were removed from the egg and transferred to petri dishes containing phosphate buffered saline (PBS) on ice. The embryos were carefully dissected using number 5 forceps, leaving only the somites and spinal cord intact. The trunk tissue was dissected to small pieces with a scalpel and transferred to a 2 ml tube in

PBS. The mixture was centrifuged at 13,200 rpm for 30 seconds and the PBS was replaced with 500 µl of lysis buffer. The tissue was homogenized by passage through a 25-gauge needle and centrifuged at 11,000 g for 20 mins at 4 ºC. The nuclei were resuspended in 150 µl of extraction buffer and incubated for 30 mins at 4 ºC before centrifugation at 21,000 g. The resulting supernatant was stored in 20 µl aliquots, which were snap frozen in liquid nitrogen and kept at -80 ºC until required.

2.3.4 – Electromobility shift assays

Binding reactions included 10 µl of 2x binding buffer, 2 µl of 1 µg/µl poly dIdC (Thermo Scientific), 3 µl of in vitro translated protein or 3 µl of nuclear extract and 2 µl of radiolabelled oligonucleotides in a final volume of 20 µl. The concentration of ficoll was reduced to 4 % in the binding buffer of reactions containing nuclear extract as it was found to greatly improve complex resolution. Supershift assays were performed by the addition of 1 µl of primary antibody to the binding reaction either before or 10 minutes after the addition of the radiolabelled probe, depending on the antibody required. Competition assays were performed by the titration of a non-labelled probe to a final 4-fold excess in the reaction.

Binding reactions were incubated on ice for 1 hour, following which the complexes were resolved on a 4 % non-denaturing polyacrylamide gel ran in 0.25x TBE. The gel was dried under vacuum at 80 ºC for 1 hour and exposed to X-ray film (Thermo Scientific) for 3 - 16 hours, depending on the strength of signal achieved. Films were developed for 2 mins using a FPM-3800A processor (FugiFilm) and digitally archived using a V750 pro scanner (Epson) linked to Photoshop (Adobe).

10x annealing buffer: 100 mM Tris pH 8.0, 10 mM EDTA and 1 M NaCl
Lysis buffer: 100 mM HEPES, 15 mM MgCl2, 100 mM KCl and 1 M DTT.
Extraction buffer: 20 mM HEPES, 1.5 mM MgCl2, 0.42 M NaCl, 0.2 mM EDTA and 25 % glycerol.
2x binding buffer: 8 % Ficoll, 40 mM HEPES, 60 mM KCL, 2 mM DTT and 0.2 mM EDTA.

2.4 – Embryo manipulation

2.4.1 – Maintenance of zebrafish lines

Wild type (WT) lines were created and maintained through pair mating of LondonAB and LondonTP stocks. Pax3 - GFP^{I150} and Pax7 - GFP^{I131} BAC transgenic stable lines were gifts from the Ingham laboratory at the University of Sheffield. These transgenic lines were maintained by outcrosses to WT and the progeny were sorted by the presence of GFP.

2.4.2 – Zebrafish embryo injection

Wild type stocks were pair mated, separated by a barrier overnight. The following morning, barriers were removed and embryos were collected every 15 minutes. The eggs were transferred to a 2.5 % agarose mould and orientated to allow easy visualization of the cell. Putative regulatory regions were injected at a concentration of 20 ng/µl in sterile water. Transposase mRNA was coinjected with all constructs contained in the MiniTol2 TKProm Gal4-UAS GFP vector at a final concentration of 14 ng/µl. Embryos were transferred to petri dishes containing aquarium water supplemented with methionine blue dye and incubated at 28.5 ˚C. Embryos were manually dechorionated and fixed with 4% paraformaldehyde (PFA) in PBS at the appropriate stage. A conversion between the embryonic stages of zebrafish development defined within Kimmel et al., 1995 to hpf of embryos maintained at 28.5 ˚C is provided in Table 2.8.

2.4.3 – Creation of transgenic zebrafish stable lines

WT embryos were injected with reporter constructs at the 1 cell stage; the resulting transient transgenics were sorted at 24 hpf on the basis of Citrine expression. Embryos were incubated 28.5 ºC until they were transferred into the care of the institute's aquatics facility at 6 days post fertilization. Injected embryos were maintained in the aquarium for approximately 4 months before being screened for germline transgenesis in outcrosses with WT stocks. Founder fish with stable germline transgene intergration were identified and maintained in single tanks. The GFP positive progeny of outcrosses between founders and WT stocks were used to expand each line

in tanks of no more than 50 fry. Pair mating of F1 fish with wildtype stocks was used to assess the activity of CNE driven trangenes.

2.4.4 – Chick *in-ovo* electroporation

Fertilized chick eggs were incubated at 37 ˚C for 40 h to allow the embryos to develop to embryonic stage 11-13 (Hamburger and Hamilton, 1951). 5 ml of albumin was removed from the egg before the shell was windowed with scissors and Indian ink was injected under the embryo to allow easy visualization and staging. Putative regulatory elements were injected at a final concentration of 500 ng/µl with transposase mRNA at 14 ng/µl, pCAGGS LacZ at 1.5 µg/µl and 5% fast green dye. DNA constructs contained in either the pCAGGS or RCAS expression vectors were injected into the lumen of the neural tube at a concentration of 1.5 – 4 µg/ml in H2O with 5% fast green dye. Platinum electrodes were placed adjacent to the embryo to deliver 50 millisecond pulses at 30 V, following which the egg was sealed with a layer of clear tape. Embryos were returned to 37 ˚C until fixation at the appropriate stage in 4 % PFA/PB.

2.5 – Embryo analysis

2.5.1 – Synthesis of riboprobes

Pax3a, *pax3b*, *pax7a* and *pax7b* antisense riboprobes were synthesised from plasmids provided by Professor Simon Hughes. Previous reports have demonstrated the specificity of these reagents for each of the zebrafish paralogous genes (Minchin and Hughes, 2008). 2 µg of plasmid DNA was digested with the appropriate restriction endonuclease (Roche and NEB) for 2 hours. Linearisation of the plamid was determined by gel electrophoresis before DNA purification using the Ilustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare). 1 µg of linear template was used in a 20 µl *in-vitro* transcription reaction containing 1x transcription buffer (Roche), 1x DIG labelling mix (Roche), 40U of RNasin (Promega) and 60U of the required RNA polymerase (Roche and Promega). The reaction was incubated at 37 ºC for 2 hours before 10U of RNase free DNase 1 (Promega) was added. The reaction was allowed to proceed for a further 15 mins to ensure all template DNA had degraded. The successful production of riboprobes was confirmed by gel electrophoresis, before the RNA was purified on a G-50 column (GE Healthcare) and stored at -20 ºC.

2.5.2 – Wholemount *in-situ* hybridization

Fixed embryos were transferred through a methanol series (25-100%) and stored at -20 ˚C for at least 24h. Embryos were rehydrated through 5 minute washes in methanol/PBS (100-25% series) followed by 4 washes in PTW at RT. Proteinase K (10 µg/ml) treatment was performed according to developmental stage, following which the embryos were re-fixed in 4 % PFA for 20 mins at RT. Embryos were washed 5x 5 mins in PTW before a brief rinse in hybridization solution followed by a 1-3h pre-hybridization step at 75 ˚C. Embryos were then incubated in hybridization solution supplemented with yeast tRNA (25 mg/ml), heparin (50 mg/ml) and the appropriate dig labeled probe (1:200 – 1:400 dilution) overnight at 75 ˚C.

The following day embryos were washed at 75 ˚C in hybridization solution/2x SSC for 20 mins, 2x SSC for 20 mins and twice in 0.2x SSC for 60 mins. RT washes with PBT/ 0.2x SSC for 10 mins and PBT for 10 mins were followed by a blocking step of 1-3h in PBT. Anti-DIG fab fragments (Roche) were diluted in PBT (1:2000) and the embryos were incubated overnight at 4 ˚C.

On the final day, the embryos were washed 4x 30 mins in PBT at RT before being transferred to a 12 well culture dish. The samples were equilibrated in staining buffer by 4x 10 mins washes before the addition of BCIP (3.5 µl/ml) and NBT (4.5 µl/ml) to the staining buffer. The culture plate was wrapped in foil to allow the staining reaction to proceed in the dark at RT for several hours. When satisfactory staining had been achieved, the embryos were washed in PTW 3x for 5 mins and fixed in 4 % PFA at 4 ˚C overnight in the dark.

2.5.3 – Zebrafish wholemount immunostaining

Fixed embryos were washed 3X 5 mins in PBT before a blocking step of 1-3h in PBDT at RT. The required primary antibodies were diluted in PBDT and the embryos were incubated at 4 ˚C overnight with constant agitation. Washes with PBT for 30 mins, repeated 4 times, were used to remove any excess primary antibody before a second blocking step of 1-3h in PBDT at RT. Alexa fluor conjugated secondary antibodies (Invitrogen) were diluted 1:400 in PBDT and the embryos were incubated at 4 ˚C overnight in the dark. On the final day the embryos were again washed in PBT 4x for 30 mins before being transferred through a glycerol series (15-75%) and stored at 4 ºC.

2.5.4 – Gelatine embedding for cryosectioning

Fixed embryos were washed 2x 5 mins in PTW before being incubated in a cryoprotective sucrose solution at 4 ºC overnight. The samples were transferred to embedding solution with a pair of forceps and incubated at 40 ºC until they had passed through the solution and settled at the bottom of the tube. The embryos were transferred to a plastic mould and orientated according to the desired plane of dissection before the solution solidified at RT. The blocks of gelatine were trimmed with a scalpel before being frozen at -42 ºC in isopentane and stored at -80 ºC until required.

2.5.5 – Immunostaining cryosections

Sections of 14 µm were prepared on a Hyrax C60 cryostat (Zeiss) and the slides were stored at -20 ˚C until required. Gelatin was removed from the samples by 5 washes in PBS at 42 ˚C before the slides were washed briefly in PBT. The tissue was blocked for 1-3 hours in PBDT at RT before the slides were incubated overnight at 4 ºC in a solution containing the required primary antibodies, diluted in PBDT. The following day, slides were washed with PBT 3x for 5 mins, blocked for 1-3h with PBDT and incubated with the appropriate secondary antibodies for 2-3h in the dark at RT. Slides were washed with PBT 3x for 5 mins, briefly dried and mounted in either Vectashield (Vector Labs) or Prolong Gold (Invitrogen) before the application of a coverslip. The antibodies used for immunohistochemistry during this investigation are listed in Table 2.7.

2.5.6 – Microscopy and image analysis

Images of *in-situ* hybridization experiments were acquired using an Axioplan 2 compound microscope equipped with an Axiocam HRC digital camera and supporting software (Zeiss). Zebrafish whole mount immunostainings were imaged using a M205FA automated stereo-microscope equipped with a DFC360FX camera and supporting software (Leica). Immunostainings on sectioned tissue were imaged using a TCS SP2 confocal microscope and supporting software (Leica). All of the acquired images were processed using both Image J and Photoshop CS4 (Adobe).

Sucrose solution: 0.12M PB with 15% sucrose

Embedding solution: 0.12M PB, 15% sucrose, 7.5% gelatine

PTW: PBS with 0.1% Tween 20

PBT: PBS, 0.1% Tween, 2% Goat serum, 0.2% BSA

PBDT: PBS, 0.1% Tween, 2% Goat serum, 0.5%BSA, 0.25 Triton-X, 0.5% DMSO

Hybridization solution: 50% Formamide, 5x SSC, 0.1% Tween 20, 9mM citric acid, dH20.

Staining buffer: 0.1M Tris pH9.5, 0.05M MgCl2, 0.1M NaCl, 0.1% Tween 20, dH20.

Table 2.1 – Primers used to clone CNEs

Name	Sequence (5' to 3')
CNE1 FW	GGGCCCAAGCTTAGTGTTGCGTTTGGCTCTTT
CNE1 RE	CCAGGTGTCATTGAGCCTCTGGACGTCCGGGCCC
CNE2 FW	GGGCCCAAGCTTAAAGCCGCAACAACAAAGAG
CNE2 RE	TCCAGTCACGCTCTAGTCCAGGACGTCCGGGCCC
CNE3 FW	GGGCCCAAGCTTTTTCTCTCCCCTCCTTCACC
CNE3 RE	AAACCGTGCATCATTTCCACGGACGTCCGGGCCC
CNE4 FW	GGGCCCAAGCTTTCACCCCCAAAAATATCTGC
CNE4 RE	TGACAATCCTCGAGCCTCATGGACGTCCGGGCCC
CNE5 FW	GGGCCCAAGCTTACTGCCTTTTCCCAGACAAA
CNE5 RE	GTTTGCAGTGGGCTGATTCTGGACGTCCGGGCCC
CNE6 FW	GGGCCCCCTGCAGGAAATCACCTCCACATTTATCGT
CNE6 RE	AGCGTTTTTGTGCCTTTTTCGGACGTCCGGGCCC
MiniTol2 3' reverse	CTGGCTAGAATCTTACTTGAG
Pax3b CNE1 FW	TTTATTGGGTCGTCTCTGTGTG
Pax3b CNE1 RE	AGCTCCTCAGGGGCAGGT
TOPO> MiniTol2 FW	GGGCCCAAGCTTTTGGTACCGAGCTCGGATCCA
TOPO> MiniTol2 RE	TCTAGATGCATGCTCGAGCGGGGACGTCCGGGCCC

Table 2.2 – Cycling parameters for mutagenic PCRs

Number of cycles	Temperature	Time
1	95ºC	1 minute
20	95ºC	50 seconds
	58ºC	50 seconds
	68ºC	10 minutes
1	68ºC	7 minutes

Table 2.3 – Primers used to delete motifs from CNEs

Name	Sequence (5' to 3')
CNE1ΔMotif1FW	GTGACTGAGCTCTTTGGAGTCTGATAGGCTCCATGAGCTGG GTCCGGG
CNE1ΔMotif1RE	CCCGGACCCAGCTCATGGAGCCTATCAGACTCCAAAGAGC TCAGTCAC
CNE1ΔMotif2FW	TGGGTCCGGGCAGCATTGAGGCCCACAATGAGGGCTGGAC GCAGGGCA
CNE1ΔMotif2RE	TGCCCTGCGTCCAGCCCTCATTGTGGGCCTCAATGCTGCCC GGACCCA
CNE1ΔMotif3FW	TTTCGGGGAGTTCTTTGGCAGTCTTAGTGAGAATGGTGACT GAGCTCT
CNE1ΔMotif3RE	AGAGCTCAGTCACCATTCTCACTAAGACTGCCAAAGAACT CCCCGAAA
CNE1ΔMotif4FW	GTAGGCTCCATGAGCTGGGTCCGGCTGTCACCTGCCTGGCA GGGGGCG
CNE1ΔMotif4RE	CGCCCCCTGCCAGGCAGGTGACAGCCGGACCCAGCTCATG GAGCCTAC
CNE3ΔMotif3FW	GCTCCCCCCACTTCAAAGATGGTTTCCTTTGAATTAATATTT GAGGATCA
CNE3ΔMotif3RE	TGATCCTCAAATATTAATTCAAAGGAAACCATCTTTGAAGT GGGGGGAGC
CNE3ΔMotif5FW	AGATGGTTTGCATTCATTAATTGTTTGAGGATCACAATGCT GGTGCCAGG
CNE3ΔMotif5RE	CCTGGCACCAGCATTGTGATCCTCAAACAATTAATGAATGC AAACCATCT

Table 2.4 – Primers used to induce mutations in CNEs

Primer Name	Sequence (5' to 3')	Target site	Mutated product
CNE1Motif1^{EboxMut1FW}	GATTTCCACGCTTTGTCCATAGGCTCCATGAG	CCATCTG	TTGTCCA
CNE1Motif1^{EboxMut1RE}	CTCATGGAGCCTATGGACAAAGCGTGGAAATC		
CNE1Motif1^{EboxMut2FW}	GATTTCCACGCTTTGTCTGTAGGCTCCATGAG	CCATCTG	TTGTCTG
CNE1Motif1^{EboxMut2RE}	CTCATGGAGCCTACAGACAAAGCGTGGAAATC		
CNE1Motif1^{EboxMut3FW}	GATTTCCACGCTCCATCCATAGGCTCCATGAG	CCATCTG	CCATCCA
CNE1Motif1^{EboxMut3RE}	CTCATGGAGCCTATGGATGGAGCGTGGAAATC		
CNE1Motif1^{RBPJMutFW}	GCTCTTTGGAGTCTGATTTCGCGCTCCATCTGTAGGCTCC	TTCCACG	TTCGGCG
CNE1Motif1^{RBPJMutRE}	GGAGCCTACAGATGGAGCGCGAAATCAGACTCCAAAGAGC		
CNE1Motif1^{TeadMut1FW}	CTGATTTCCACGCTCCATCTGTATTCTCCATGAGCTGGGTCCGGG	TAGGCT	TATTCT
CNE1Motif1^{TeadMut1RE}	CCCGGACCCAGCTCATGGAGAATACAGATGGAGCGTGGAAATCAG		
CNE1Motif1^{TeadMut2FW}	CCACGCTCCATCTGTAGGCTAAATGAGCTGGGTCCGGGCAGC	CTCCAT	CTAAAT
CNE1Motif1^{TeadMut2RE}	GCTGCCCGGACCCAGCTCATTTAGCCTACAGATGGAGCGTGG		
CNE1Motif3^{Mut1FW}	GAGTTCTTTGGCAGTCTATAGTTGGCCTGTTTAGTGAGAA	TTTGTT	ATAGTT
CNE1Motif3^{Mut1RE}	TTCTCACTAAACAGGCCAACTATAGACTGCCAAAGAACTC		

CNE1Motif3^{Mut2FW}	GAGTTCTTTGGCAGTCTTTTG CCGGCCTGTTTAGTGAGAA	TTTGTT	TTTGCC
CNE1Motif3^{Mut2RE}	TTCTCACTAAACAGGCCGGC AAAAGACTGCCAAAGAACT C		
CNE3Motif3^{HoxMutFW}	AGATGGTTTGCATTCATTCC TTGTCCTTTGAATTAATA	TTAATT	TTCCTT
CNE3Motif3^{HoxMutRE}	TATTAATTCAAAGGACAAGG AATGAATGCAAACCATCT		
CNE3Motif3^{FoxMutFW}	AAGATGGTTTGCATTCACCC ATTGTCCTTTGAATTAAT	CATTAAT	CACCCAT
CNE3Motif3^{FoxMutRE}	ATTAATTCAAAGGACAATGG GTGAATGCAAACCATCTT		
CNE3Motif5^{HoxMutFW}	CATTAATTGTCCTTTGAATTC CTATTTGAGGATCACAATGC	TAATA	TCCTA
CNE3Motif5^{HoxMutRE}	GCATTGTGATCCTCAAATAG GAATTCAAAGGACAATTAA		

Table 2.5 – Constructs used for *in-vitro* protein synthesis

cDNA	Source	Vector	RNA Polymerase
Hoxa5	ImaGenes	pCMV-SPORT6	Sp6
Hoxa7	Imagenes	pCMV-SPORT6	Sp6
Lef1	Imagenes	pCMV-SPORT6	Sp6
Pax3	ImaGenes	pCR-Blunt*II-TOPO*	Sp6
Pax7	ImaGenes	pCR-Blunt*II-TOPO*	Sp6
Pbx1	ImaGenes	pYX-Asc	T7
Six2	Dr Eva Kutejova	PcDNA3	T7
Sox2	Lovell-Badge lab, NIMR	pCI-NEO	T7
Tcf3	ImaGenes	pCMV-SPORT6	Sp6
Tcf7	ImaGenes	pCMV-SPORT6	Sp6
Tead2	ImaGenes	pSPORT1	T7

Table 2.6 – Oligonucleotides used as EMSA probes

Oligo name	Sequence (5'to 3')
CNE1Motif1FW	GAGTCTGATTTCCACGCTCCATCTGTAG GCTCC
CNE1Motif1RE	GGAGCCTACAGATGGAGCGTGGAAATC AGACTC
CNE1Motif1$^{E\text{-}boxMut3FW}$	GATTTCCACGCTCCATCCATAGGCTCCA TGAG
CNE1Motif1$^{E\text{-}boxMut3RE}$	CTCATGGAGCCTATGGATGGAGCGTGGA AATC
CNE1Motif3FW	GGCAGTCTTTTGTTGGCCTGTTTAGTGA GAATG
CNE1Motif3RE	CATTCTCACTAAACAGGCCAACAAAAGA CTGCC
CNE3Motif3-5FW	CCACTTCAAAGATGGTTTGCATTCATTA ATTGTCCTTTGAATTAATAT
CNE3Motif3-5RE	ATATTAATTCAAAGGACAATTAATGAAT GCAAACCATCTTTGAAGTGG
CNE3Motif3^{HoxMutFW}	AGATGGTTTGCATTCATTCCTTGTCCTTT GAATTAATA
CNE3Motif3^{HoxMutRE}	TATTAATTCAAAGGACAAGGAATGAATG CAAACCATCT
CNE3Motif3^{FoxMutFW}	AAGATGGTTTGCATTCACCCATTGTCCTT TGAATTAAT
CNE3Motif3^{FoxMutRE}	ATTAATTCAAAGGACAATGGGTGAATGC AAACCATCTT
CNE3Motif5^{HoxMutFW}	CATTAATTGTCCTTTGAATTCCTATTTGA GGATCACAATGC
CNE3Motif5^{HoxMutRE}	GCATTGTGATCCTCAAATAGGAATTCAA AGGACAATTAA

Table 2.7 – Primary antibodies used for immunohistochemistry

Antibody	Species	Origin	Dilution
B-Galactosidase	Rabbit	ABD Serotec (AHP1292)	1:500
ChAT	Rabbit	Gift from Dr. Vassilis Pachnis	1:1000
DP311	Mouse	Gift from Dr. Nipam Patel	1:75
DP312	Mouse	Gift from Dr. Nipam Patel	1:75
GFP	Rabbit	Molecular Probes (A11122)	1:1000
GFP	Sheep	Biogenesis	1:1000
Olig2	Goat	R&D Systems (AF2418)	1:1000
Pax3	Mouse	DSHB	1:20
Pax6	Mouse	DSHB	1:20
Pax7	Mouse	DSHB	1:20
Rhodamine conjugated Phalloidin	N/A	Molecular Probes (R415)	1:500

Table 2.8 – Conversion of zebrafish developmental stages to hpf

Developmental stage	hpf at 28.5 ºC
Bud	10
6 somites	12
12 somites	15
18 somites	18
Prim-5	24
Prim-15	32
High-pec	48
Protruding mouth	72

Chapter 3: The establishment of the Pax3 domain in the neural tube

As described in chapter 1.3 ,the anatomical division of the neural tube by the sulcus limitans is commonly accepted to mark the division of spinal cord progenitors into sensory and motor fates. We aimed to investigate this hypothesis by using transgenic mice to genetically trace the progeny of the alar plate from the early stages of neural tube patterning to the establishment of functional neuronal circuits within the spinal cord at birth. We chose to utilise Pax3 as a molecular marker of the alar plate in these studies, as previous reports have demonstrated that it is expressed in all progenitors of the dorsal spinal cord and shares a ventral boundary with the sulcus limitans between E10-13 in mice (Goulding et al., 1991). We reasoned that the initiation of Pax3 transcription within the closing neural folds also provided an opportunity to assess progenitor dispersal and the establishment of the Pax3 expression domain before the formation of the sulcus limitans.

3.1 – *Pax3Cre/Rosa26-YFP* mice as a tool for lineage tracing

Pax3-Cre transgenic mice were obtained as part of a collaboration we initiated with the laboratory of Dr Fredéric Relaix. These mice were created by targeted replacement of the first exon of *Pax3* with Cre recombinase by homologous recombination and therefore express Cre under the control of the endogenous *Pax3* regulatory elements (Engleka et al., 2005). The transgenic allele does not produce a functional Pax3 protein and homozygous $Pax3^{Cre/Cre}$ mice exhibit severe neural tube, neural crest and musculature defects that are comparable to those observed in homozygous *splotch* mutants (Engleka et al., 2005). Heterozygous $Pax3^{Cre/+}$ embryos are viable and develop normally, with the exception of a mild pigment defect that is produced as a result of reduced Pax3 protein levels. *Pax3-Cre* transgenics were crossed with mice in which an inducible yellow fluorescent protein (YFP) transgene, localised to the cell membrane, is targeted to the Rosa26 locus. This genomic location is commonly used in lineage studies as it has been shown to allow unrestricted expression of reporter genes (Friedrich and Soriano. 1991; Zambrowicz et al., 1997). The expression of YFP from the Rosa26 locus is dependent upon the removal of a floxed stop cassette by Cre recombinase, resulting in the persistent expression of YFP in cells that have expressed Pax3 at any point during their development (Fig. 3.1). $Pax3^{Cre/+}$*/Rosa26-YFP* embryos were generated by members of the Relaix lab, collected

at the desired stage and genotyped before further analyses were independantly performed by both laboratories.

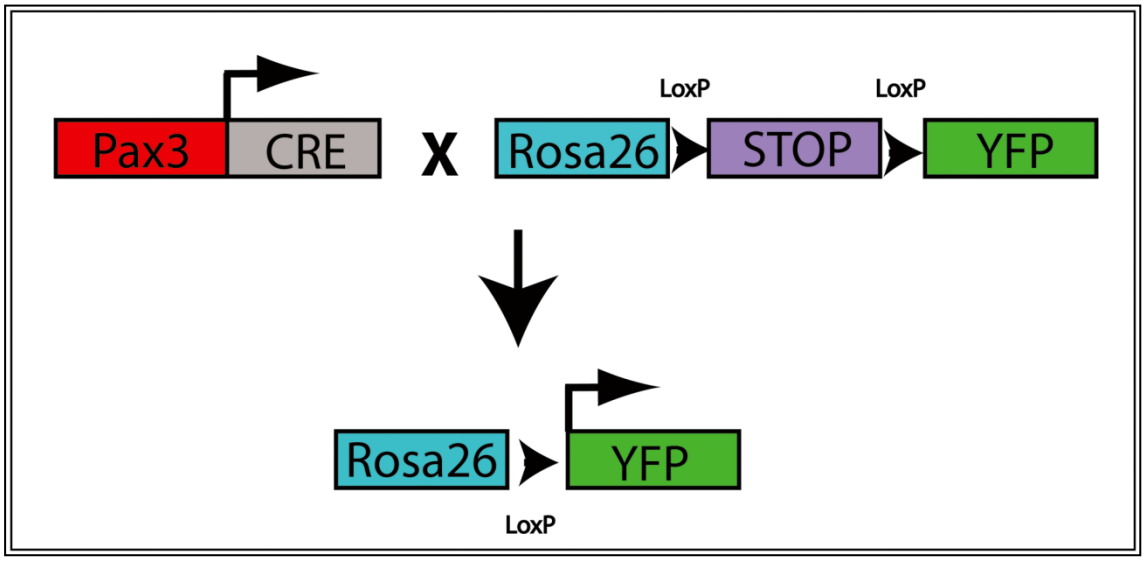


Fig. 3.1 – Genetic strategy for lineage tracing Pax3 expressing progenitors

Mice expressing Cre recombinase under the control of Pax3 regulatory elements were crossed with Rosa26 YFP transgenics. In Pax3 expressing cells, Cre mediates a recombination event that results in the excision of the stop cassette from the Rosa26 locus and the continued expression of YFP in Pax3 progeny.

3.2 – The Pax3 domain is established by a switch in progenitor identity

Pax3 transcription is initiated in the neural plate at E8.5 (Goulding et al., 1991), however the earliest stage assessed during this study was E9.5. Anterior regions of the spinal cord had completed neurulation by this stage but the sulcus limitans was not yet clearly defined, which permitted the analysis of progenitor dispersal in the neuroepithelium prior to the establishment of the alar plate. Pax3 antibody staining at the hindbrain level was restricted to the dorsal half of the neural tube (n=2) (Fig. 3.2A). The GFP labelled Pax3 lineage was also largely restricted to the neural tube, however we noted that YFP expressing cells were present beyond the ventral limit of the Pax3 expression domain (Fig. 3.2B). Analysis of brachial and thoracic sections demonstrated that Pax3 was expressed within the dorsal spinal cord and the dermomyotome (Figs. 3.2C and E), however cells marked by the presence of YFP were also found throughout the developing sclerotome (Figs. 3.2D and F). We observed two distinct populations of YFP labelled cells that resided ventrally to the Pax3 domain within the spinal cord. The first reproducibly spanned 3-4 cells in the DV axis and was located adjacent to the Pax3 ventral boundary (Arrows in Figs. 3.2D and F), in agreement with our observations at hindbrain level. We also observed isolated clones of YFP labelled cells that were dispersed in the ventral domains of the spinal cord (Asterisks. in Figs 3.2D and F). The distribution of these clones in the neuroepithlium varied both between siblings and along the anterior-posterior axis of individual embryos.

These data were supported by an independent analysis of the Pax3 lineage, performed by Dr Vanessa Ribes in the Relaix lab at E9.75. These experiments employed a mouse strain in which the Rosa26 locus was targeted with both Tomato and GFP reporter genes but used the same *Pax3-Cre* driver as our study. Her results indicated that Pax3 lineage extended beyond the ventral boundary of gene expression in the spinal cord (n=3). These data also indicated that GFP labelled cells were located both adjacent to the Pax3 expression domain and in clones within the ventral spinal cord. These data suggested that the Pax3 lineage was not a molecular correlate of the embryonic alar plate, due to a transcriptional switch in cell identity that occured before E9.5 in mice.

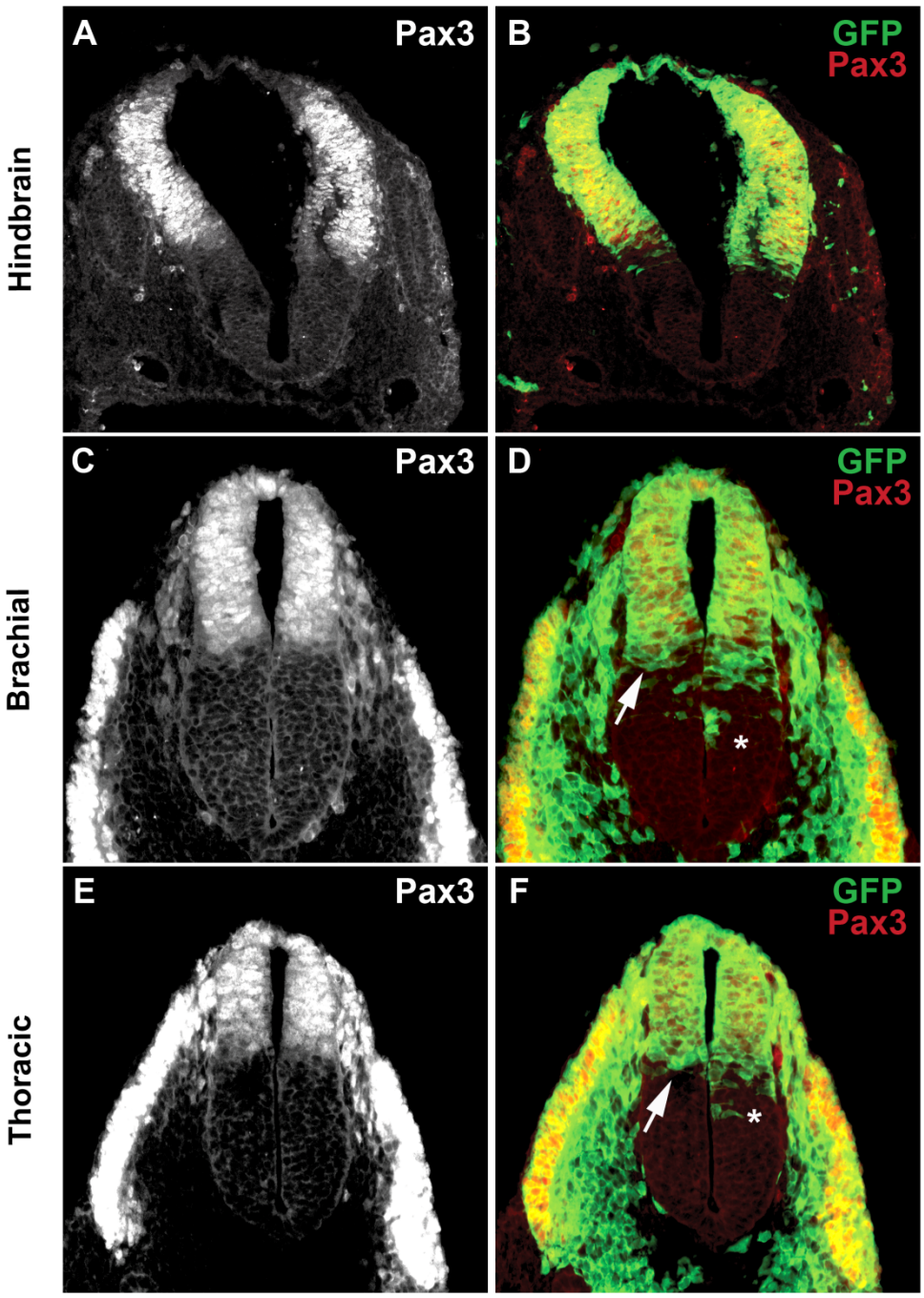


Fig. 3.2 – The position of the Pax3 domain is established before E9.5

Pax3 protein is expressed in the dorsal neural tube at hindbrain (A), brachial (C) and thoracic levels (E) at E9.5. YFP labelling, visualised with a polyclonal GFP antibody, is observed beyond the ventral boundary of Pax3 expression at all AP levels assessed (B, D and F). $Pax3^{-}$ YFP^{+} cells are distributed within two populations, a domain of cells adjacent to the ventral boundary of Pax3 expression (Arrows in D and E) and as isolated ventral clones (Asterisks in D and E). Cells marked by YFP ptotein are also observed throughout the somites and sclerotome surrounding the neural tube, indicating that Pax3 is transiently expressed in all mesoderm progenitors.

3.3 – The position of the Pax3 domain remains constant during development

The embryonic alar plate is defined in the spinal cord by the formation of the sulcus limitans at approximately E10.5 (Goulding et al., 1991 ; Gross et al., 2002). We hypothesised that this anatomical division of the neural tube may restrict cell migration and contribute to the maintenance of the position of the ventral Pax3 expression boundary within the tissue. Consistent with our previous observations, Pax3 protein expression was restricted to the dermomyotome at E10.5 (Figs. 3.3A and C) whereas its YFP labelled lineage was distributed throughout the surrounding mesoderm (n=2) (Figs. 3.3B and D). Within the neural tube, GFP staining was reproducibly observed adjacent to the ventral boundary of the Pax3 domain along the anterior-posterior axis of both embryos assessed at E10.5 (Arrows in Figs. 3.3B and D). Clones of YFP labelled cells were also evident in the ventral domains of the spinal cord, supporting our earlier observations (Asterisks in Figs. 3.3B and D). We investigated the dispersal of these cells within the neuroepithelium in relation to the expression profiles of well-characterised ventral patterning genes. These studies revealed the presence of GFP labelled cells throughout the Pax6 domain of the intermediate spinal cord (Figs. 3.4A and B) and isolated clones within the pMN domain marked by Olig2 expression (Figs. 3.4C and D).

Our observations of the Pax3 lineage at E10.5 were supported by independent experiments performed by the Relaix lab at E11.75. Sections *Pax3*$^{Cre/+}$*/Rosa26-Tomato/GFP* embryos exhibited the downregulation of Pax3 expression within the dorsal spinal cord that occurs during the process of neurogenesis (n=2) (Figs. 3.5A, C and E). This was clearly visible as GFP staining within Pax3 derived neurons in the mantle zone (Figs. 3.5B,D and F) and the axons of dI1 and dI2 interneurons that crossed the midline at the floor plate (Figs. 3.5B, D and F) (Gross et al., 2002 ; Helms and Johnson., 1998). In accordance with our data, the ventral boundary of Pax3 expression in the neural tube was located adjacent to a clearly defined domain of cells labelled solely by GFP staining (Figs. 3.5B, D and F). Dr Ribes also sought to determine the location of these respecified cells relative to V0 interneurons, which are marked by the expression of Evx1 (Moran-Rivard et al., 2001). These data demonstrated that all Evx1 expressing interneurons are derived from cells that have expressed Pax3 during their early development (Figs. 3.6A, B and C). Furthermore, she also observed the presence of YFP labelled clones dispersed throughout the ventral domains of both embryos assessed at E11.75 (Figs. 3.5B, D and F).

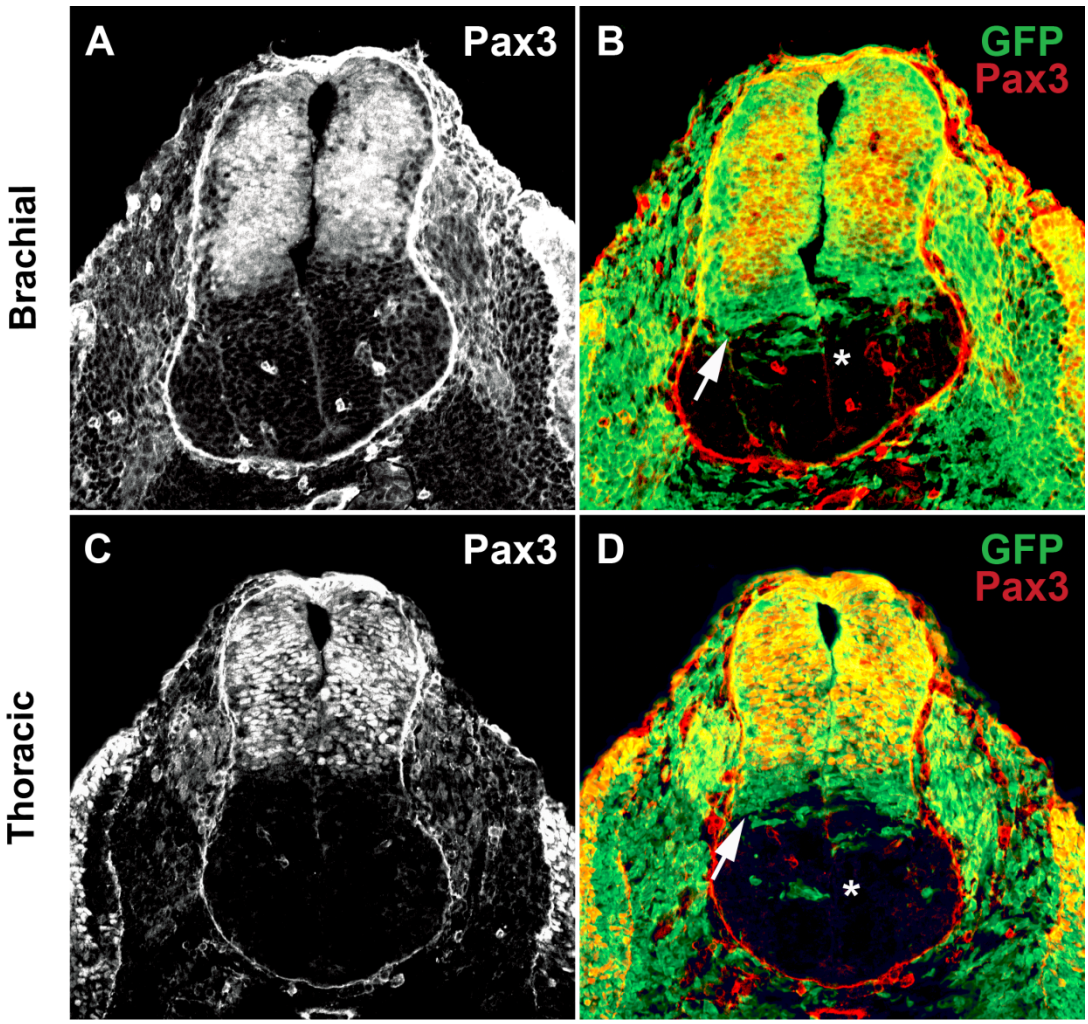


Fig. 3.3 – The distribution of the Pax3 lineage during neural tube patterning

Pax3 expression is restricted to the dorsal neural tube at brachial (A) and thoracic (C) levels at E10.5. GFP antibody staining labells a domain of cells adjacent to the boundary of Pax3 expression (Arrows in B and D) and isolated clones within the ventral domains of the neural tube (Asterisks in B and D), consistent with data obtained at E9.5 (Fig. 3.2).

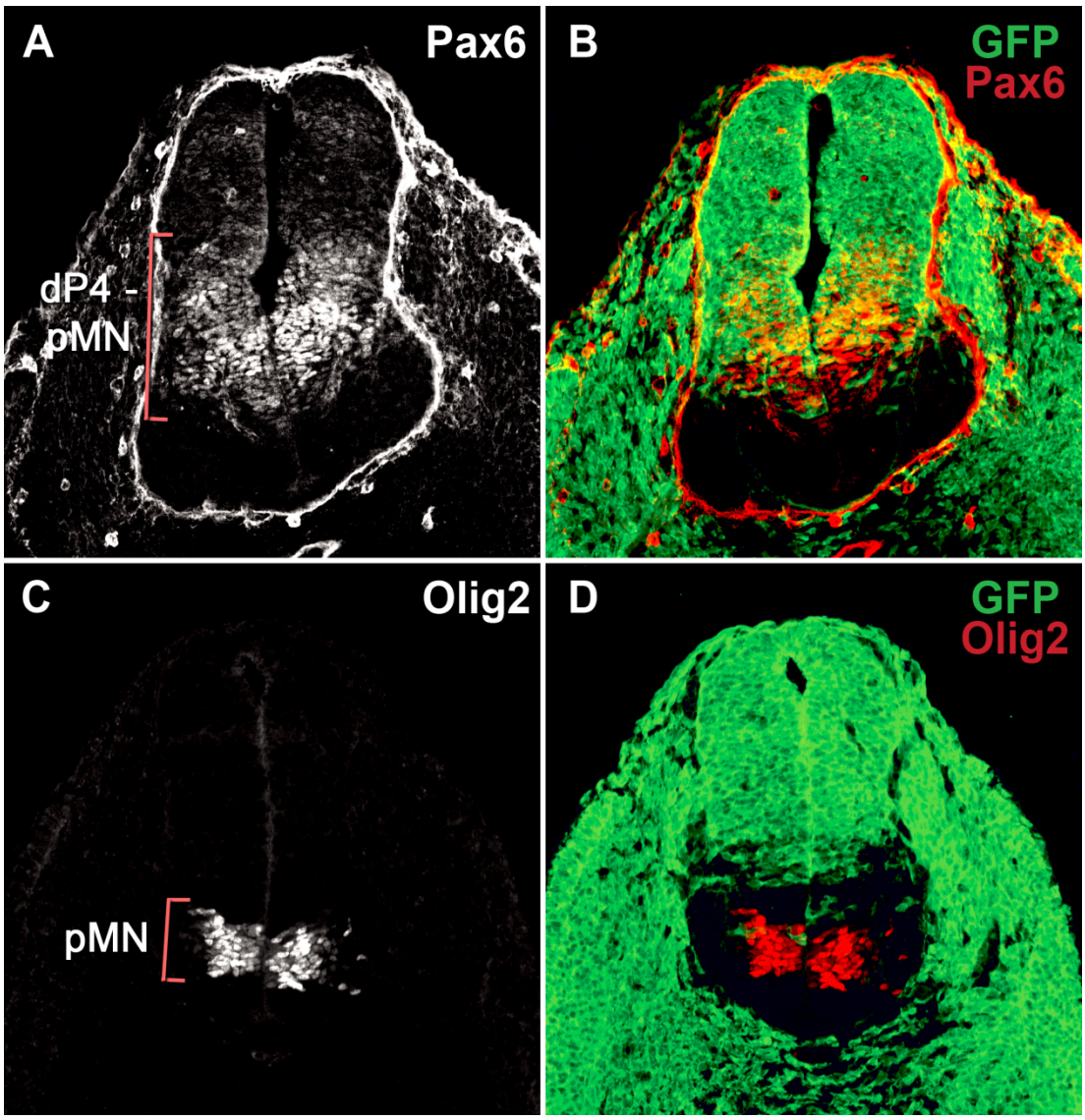


Fig. 3.4 – Clonal dispersion of Pax3 progeny within the ventral neural tube

Brachial sections taken at E10.5 reveals Pax6 expression from the dP4 domain dorsally to the pMN domain in the ventral neural tube (A). YFP labelling is observed in cells located throughout the Pax6 expression domain of the intermediate neural tube (B). Olig2 specifically labels motorneuron progenitors within the pMN domain of the ventral neural tube (C). Isolated clones of YFP labelled cells are infrequently observed within the pMN domain (D).

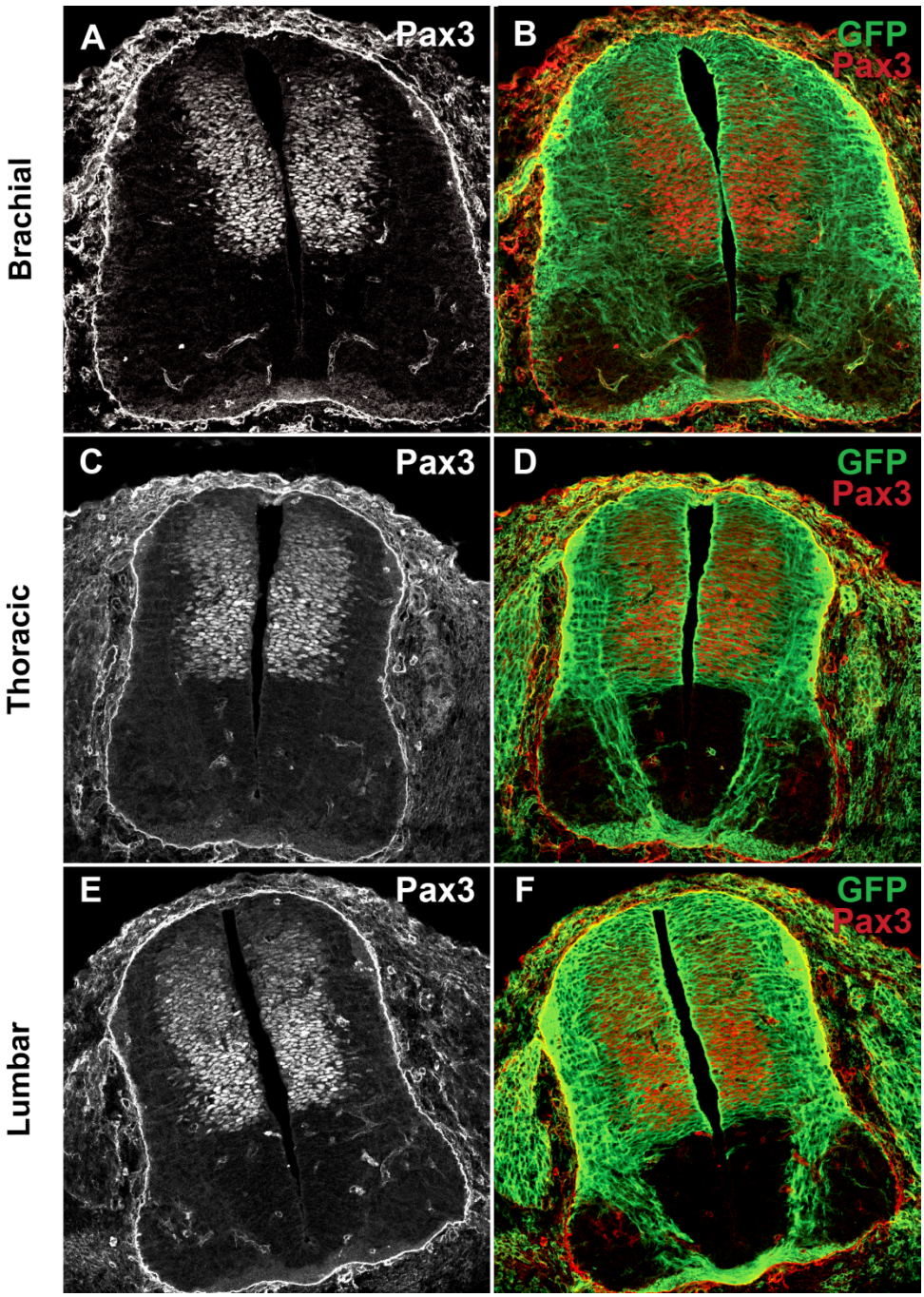


Fig. 3.5 – The distribution of the Pax3 lineage during early neurogenesis

Pax3 is expressed in progenitors within the ventricular zone of the alar plate at brachial (A), thoracic (C) and lumbar (E) levels at E11.75. GFP labelled cells are present in the ventral spinal cord as both a continuous domain adjacent to the Pax3 ventral boundary and as isolated clones (B, D and E). GFP staining is also evident in postmitotic neurons within the mantle zone of the alar plate and the axons of commissural interneurons that project ventrally towards the FP. Data kindly provided by Dr. Vanessa Ribes.

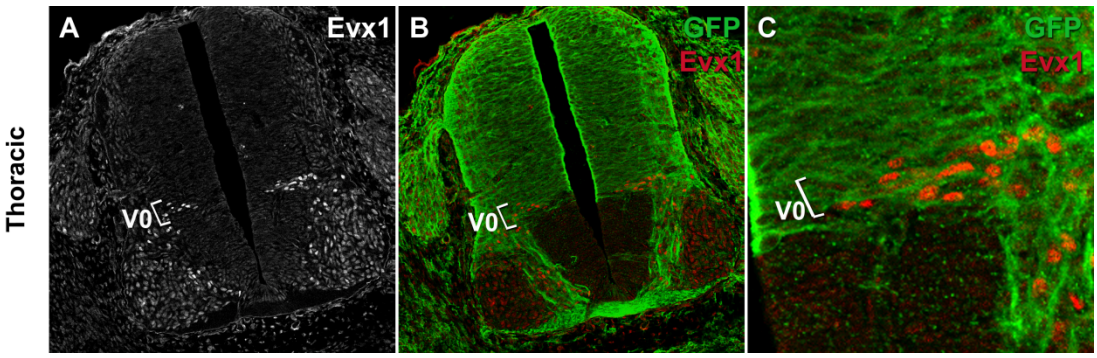


Fig. 3.6 – V0 interneurons transiently express Pax3 during their development

Evx1 antibody staining labels progenitor cells committed to a V0 interneuron fate (A). GFP staining at E11.75 reveals that all Evx1 expressing progenitor cells are derived from cells that have previously expressed Pax3 (B and C).

Data kindly provided by Dr. Vanessa Ribes.

3.4 – The dispersal of Pax3 progeny in the functional spinal cord

Highly coordinated programs of neurogenesis, migration and synaptogenesis establish functional sensory-motor connections within the spinal cord by birth, also known as postnatal day (P)0 of mouse development (Ladle et al., 2007). We sought to determine the fate of Pax3 expressing progenitors within the spinal cord at P0 to investigate any restriction of this lineage to sensory interneurons. Our previous analyses demonstrated that the Pax3 lineage encompassed, but was not restricted to, the embryonic alar plate. Consequently, these experiments were not able to directly assess any restriction of cell fate that may be imposed by the sulcus limitans at later stages of neural tube development. However, we reasoned that analysis of the Pax3 lineage at P0 might still prove to be informative if interpreted with the knowledge that V0 interneurons were reproducibly labelled in these samples.

Transverse sections of *Pax3*$^{Cre/+}$*/Rosa26-YFP* embryos prepared at the brachial and thoracic levels of the spinal cord revealed uniform YFP labelling throughout the grey and white matter located dorsally to the central canal (n=2) (Figs. 3.7A and C). Extensive GFP labelling was observed within the lateral and ventral fassiculi of white matter tracts that surround the ventral horn of grey matter (Figs. 3.7A and B). YFP labelled neurons were observed medially within the ventral horn, consistent with the position of V0 interneurons in the tissue following their ventral migration (Figs 3.7B and D) (Lanuza et al., 2004 ; Moran-Rivard et al., 2001 ; Pierani et al., 2001). We reasoned that these labelled cells may also represent populations of dI3 and dI6 interneurons that migrate into the ventral horn and are thought to provide direct sensory input to motor circuits (Avraham et al., 2010 ; Gross et al., 2002).

We did not observe GFP staining in the lateral regions of the motor horns at P0, indicating that these cells had not expressed Pax3 at any stage of their development. We hypothesised that these regions might correspond to the location of the basal plate derived V1, V2, motor and V3 classes of neurons within the tissue. This point was investigated in experiments determining the localisation of YFP labelled cells relative to choline acetyltransferase (ChAT) immunostaining. This enzyme is required for the production of acetylcholine and is frequently used as a specific marker of motor neurons within the spinal cord (Tanabe et al., 1998 ; Arber et al., 1999). Analysis of thoracic sections revealed the presence of YFP labelled neurons within the medial region of the ventral horn as previously described (Fig. 3.8A), whereas ChAT labelled motor neurons were localised at the lateral limit of the grey matter (Fig. 3.8B). These two cell

populations appeared to be mutually exclusive, however we frequently observed YFP labelled axons both adjacent to and within pools of motor neurons (Fig. 3.8C). Identification of the Pax3 derived cells involved in these in these sensory-motor interactions was not possible in this investigation because Pax3 transcription is initiated in multiple classes of interneurons that are known to synapse with motor neurons. Many cells within the ventral horn were not labelled by GFP or ChAT antibody staining, consistent with our previously described data which indicated that only V0 interneuron progenitors were reproducibly labelled by the Pax3 lineage within the basal plate (Figs 3.5B,D and F).

The data produced from our lineage study suggest that Pax3 expression is not a molecular correlate of the embryonic alar plate in mouse embryos, due to the initiation of transcription within the neural plate and the dynamic changes in cell identity that occur within this tissue. We observe two distinct classes of cells that appear to extinguish Pax3 expression during their development. The first are randomly dispersed throughout the ventral neural tube as clones, in agreement with reports of extensive cell mixing across the Pax3 boundary at early stages of chick neural development (Erskine et al., 1998 ; Leber and Sanes, 1995). Our data also demonstrate that a domain of Pax3 labelled cells is respecified to a more ventral p0 progenitor identity, this result is consistent both between embryos and along the DV axis of the neural tube. We hypothesise that this transcriptional switch in cell identity positions the Pax3 expression domain in the neural tube before E9.5. Once established, the position of the domain in the tissue appeared to remain constant in our study, possibly due to restricted migration enforced by cell adhesion molecules expressed at the sulcus limitans. Our analysis of *Pax3*$^{Cre/+}$*/Rosa26-YFP* mice at P0 provide further evidence to support this conclusion, as we demonstrated that the Pax3 lineage does not give rise to neither motor neurons nor clusters of interneurons within the ventral horn.

This study raises questions concerning the regulatory logic that establishes the Pax3 expression domain during the early stages of neurulation and maintains its position in the neural tube during CNS development. Our results clearly demonstrate that all cells located within the alar plate are derived from the Pax3 lineage, whereas the majority of basal plate progenitors do not express the gene at any point in their development. These data indicate that regulatory mechanisms operate at the earliest stages of CNS development to initiate Pax3 expression within a restricted population of progenitor cells within the neural plate. As discussed in Chapter1.4, the molecular events that initiate Pax3 transcription within the neural plate are currently poorly

understood. Our studies also revealed transient Pax3 expression within a population of ventral interneuron progenitors. These observations correlate with published data demonstrating that Pax3 expression is repressed by Shh signaling and the ventral patterning genes it induces; however the mechanism of this repressive interaction is unknown. We sought to investigate the regulatory logic that governs Pax3 expression during CNS development by identifying functional enhancers associated with this locus and determining the molecular interactions between these elements and members of the neural tube GRN. We also aimed to compare the function of enhancers associated with the Pax3 and Pax7 loci in order to investigate the basis of their differing spatiotemporal expression profiles.

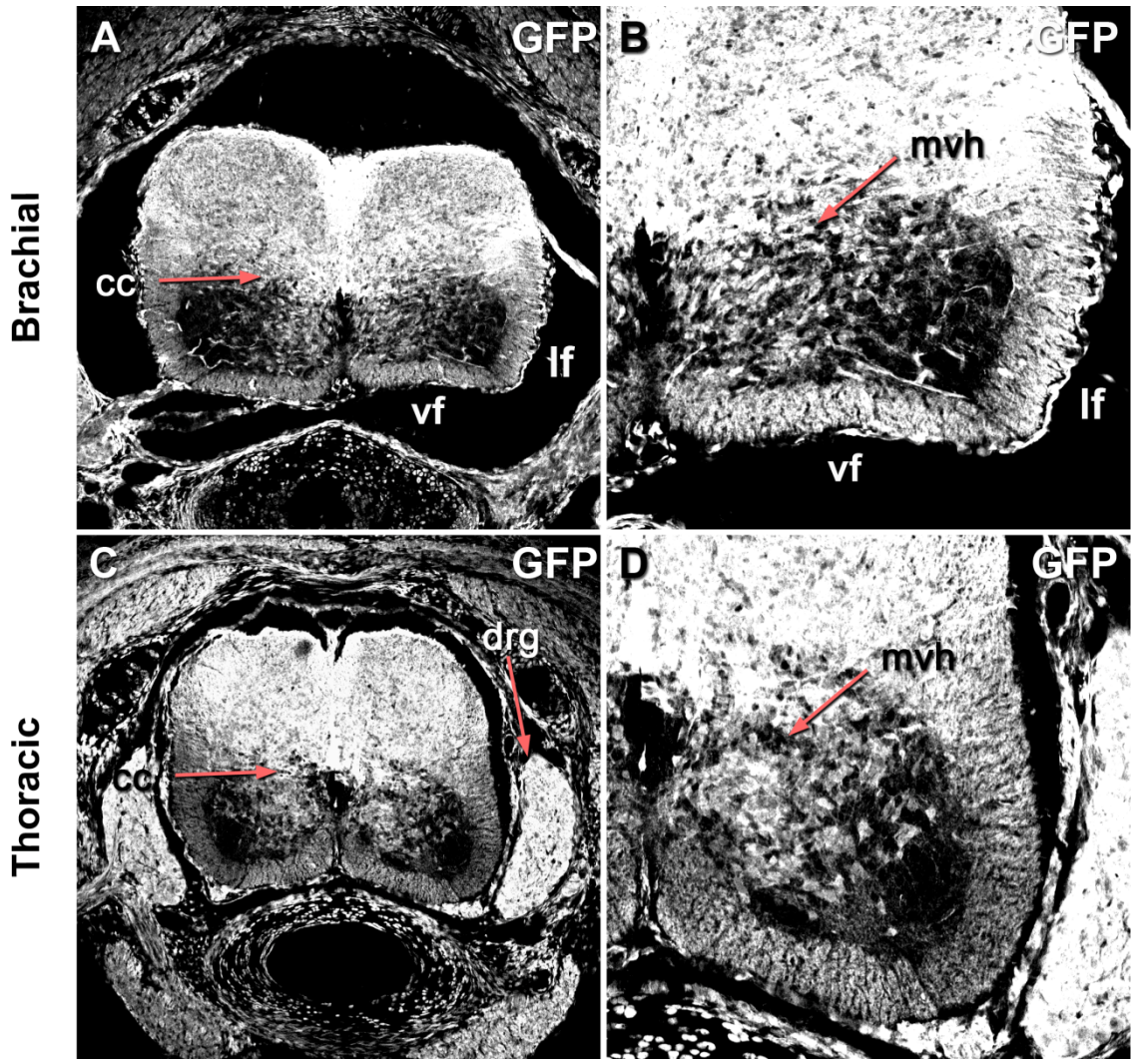


Fig. 3.7 – The distribution of the Pax3 lineage at birth

YFP labelling is observed throughout the grey and white matter located dorsally to the central canal (cc), consistent with the expression of Pax3 in the embryonic alar plate (A and C). The lateral and ventral funiculi of white matter tracts surrounding the ventral horn are labelled by YFP protein (lf and vf respectively). Populations of V0, dI3 and dI6 interneurons located within the medial region of the ventral horn (mvh) are derived from Pax3 expressing progenitors (B and D). GFP staining is also observed throughout the dorsal root ganglion (drg), consistent with the expression of Pax3 in neural crest derivatives.

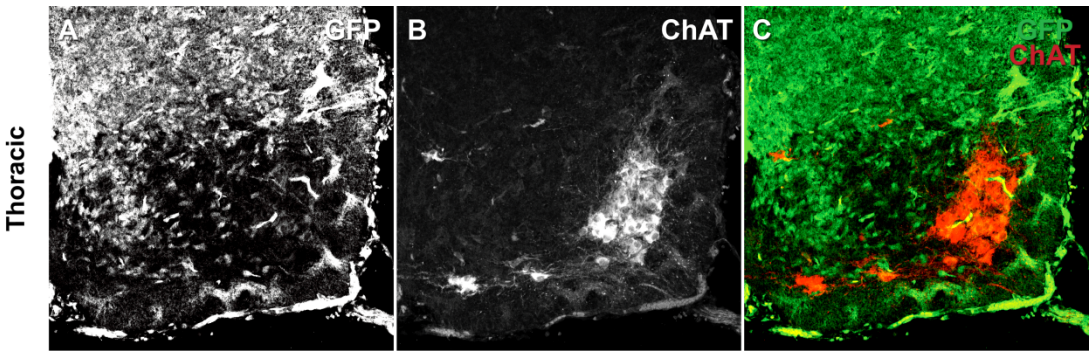


Fig. 3.8 – The position of the Pax3 domain is maintained throughout development

YFP labelling is restricted to interneurons within the medial domain of the ventral horn (A), whereas ChAT stained motor neurons are located laterally (B). Co-localisation of YFP and ChAT labelling reveals that the Pax3 lineage is segregated from motor neuron progenitors from the earliest stages of neural plate patterning (C). Populations of cells within the ventral horn are not labelled by either GFP or ChAT antibody staining, suggesting that they might represent basal plate derived interneurons (C).

Chapter 4: Establishing Zebrafish as a model for studying Pax3/7 regulation

In order to investigate the molecular basis of Pax3/7 regulation, it was first necessary to select a model system in which the activity of putative regulatory elements could be assessed *in-vivo*. Zebrafish are increasingly used in functional screens of putative enhancers due to the availability of large numbers of optically transparent embryos, their comparatively rapid rate of development and a sequenced genome (Gomez-Skarmeta et al., 2006 ; Müller et al., 2002). These advantages were the major factors in the selection of zebrafish embryos as our *in-vivo* system for screening putative enhancers.

The cellular mechanism of neurulation in zebrafish embryos differs from that observed in mice and chick, as the neural plate does not fold to give rise to the neural tube. Instead, cells within the zebrafish neural plate converge at the midline to give rise to the neural keel, which subsequently cavitates to form the neural tube (Kimmel et al., 1994 ; Schmitz et al., 1993). Nevertheless, the fundamental principles of neural tube patterning are conserved between zebrafish and higher vertebrates (reviewed by Appel, 2000 ; Lewis and Eisen, 2003 and Wilson et al., 2002). In agreement with these data, Pax3/7 expression in zebrafish embryos has been shown to be repressed by exogenous shh ligand (Guner and Karlstrom. 2007). However in contrast to higher vertebrate models, a ventral expansion of these genes is not observed in an array of shh pathway mutants (Guner and Karlstrom. 2007). Furthermore, *pax3* expression has been shown to be repressed by the ectopic induction of the Wnt pathway agonist *dickkopf1*, morpholino knockdown of *wnt8a* or loss of tcf/lef activity (Bonner et al., 2008 ; Garnett et al., 2012). Taken together, these data demonstrate that the key concepts of both spinal cord patterning and Pax3/7 regulation are conserved in zebrafish embryos, validating them as a potential model in which to base our study.

The zebrafish genome contains two parologous genes for both Pax3 and Pax7, as a result of a whole genome duplication in the teleost lineage (Postlethwait et al., 1998). It was first necessary to define the endogenous expression patterns of each of these genes to establish zebrafish embryos as a viable model in which to perform a screen of putative regulatory elements. This was achieved through a series of *in-situ* hybridisation experiments and literature searches, which determined the expression of each of the Pax3/7 paralogs during the first 48 hours of zebrafish development.

4.1 – Pax3 paralogs are first expressed in the early neural plate

Pax3a transcripts were first observed in the intermediate and lateral regions of the posterior neural plate at bud stage (Fig. 4.1A) (Table 2.8). Stripes of medio-lateral expression corresponding to the presumptive midbrain and hindbrain were clearly visible in the anterior neural plate at this stage. Expression became restricted to the most lateral regions of the neural plate by the 6 somites stage and the rostral stripes of transcripts were fused at the midline, defining the developing midbrain and rhombomeres 2 and 4 of the hindbrain. This expression profile was maintained during early segmentation stages until the formation of the neural tube at trunk level (Data not shown). As neurulation was nearing completion at the 18 somites stage, *pax3a* mRNA was restricted to the dorsal half of the neural tube (Fig. 4.1C). Transcripts were detected across the entire anterior-posterior axis of the CNS, with the exception of the developing forebrain. This expression profile was maintained at 24 hpf, at which stage transverse sections of trunk tissue clearly demonstrated that *pax3a* mRNA was restricted to the dorsal half of the zebrafish spinal cord (Figs. 4.1D and 4.3A). *Pax3a* was also observed in the newly formed somites within the posterior trunk region at 24 hpf, consistent with its critical functions during myogenesis (Fig. 4.1D). Transcription within the spinal cord appeared to be decreased by 32 hpf as mRNA was localised within only the most dorsally located cells (Fig. 4.1E). Anteriorly, the dorsal tectum, midbrain and hindbrain expressed *pax3a* at this stage. *Pax3a* mRNA was not detected in the spinal cord at 48 hpf by *in-situ* hybridisation, however strong expression persisted in the dorsal hindbrain, midbrain and tectum (Fig. 4.1F).

Several attempts were made to assess the expression profile of *pax3b* during CNS development by *in-situ* hybridization, however the quality of the results obtained was unsatisfactory and made it difficult to draw any meaningful conclusions (Data not shown). A previous report has indicated that *pax3b* is expressed in the CNS during zebrafish development (Minchin and Hughes, 2008). In this study, expression of *pax3b* was first reported in the presumptive midbrain and hindbrain at the 7 somites stage. These domains of expression appeared to correlate well with cells expressing *pax3a* in stage matched embryos. Expression of *pax3b* did not extend to the posterior regions of the embryo and appeared to be restricted to the lateral regions of the anterior neural plate. *Pax3b* mRNA was observed in the midbrain and hindbrain at 25 hpf, however the anterior limit of expression appeared to be reduced in comparison to *pax3a*. Sections of

spinal cord tissue suggested that *pax3b* mRNA was dorsally expressed but did not extend to the ventral limit of *pax3a* expression.

Together, these data demonstrate that the expression profiles of the zebrafish Pax3 paralogs closely correlates with that observed in chick and mouse embryos (Goulding et al., 1991 ; Goulding et al 1993), validating zebrafish embryos as a model system in which to base our study. As in other species, transcription of Pax3 is first observed early during neural tube development in the intermediate and lateral regions of the open neural plate. Upon completion of neurulation, *pax3a* expression appears to define the dorsal spinal cord, whereas *pax3b* appears to exhibit a more restricted profile.

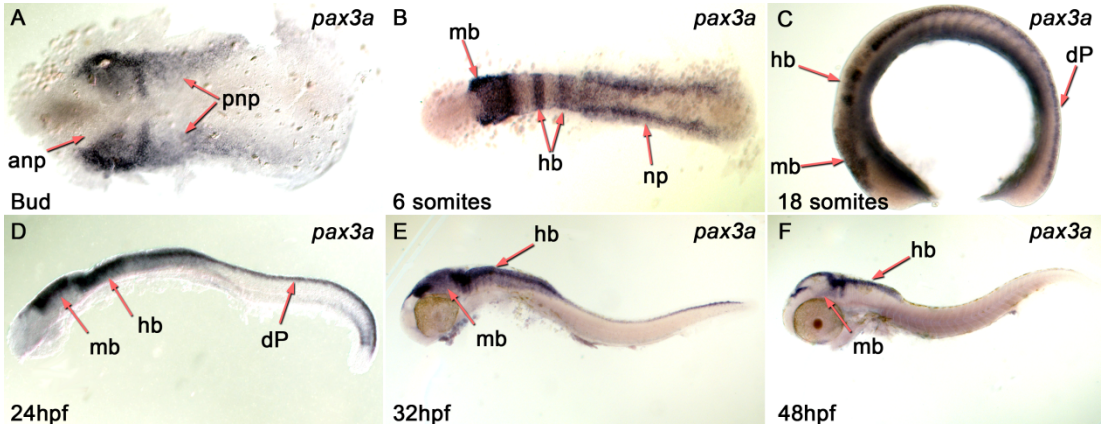


Fig. 4.1 – The profile of *pax3a* expression during zebrafish CNS development

Pax3a transcripts are first detected at bud stage in medio-lateral stripes within the anterior neural plate (anp) and within the intermediate regions of the posterior neural plate (pnp) (A). By 6 somites, the anterior stripes fuse to define the midbrain (mb) and rhombomeres 2 and 4 of the hindbrain (hb) ; expression is restricted to the lateral limit of the posterior neural plate (np) (B). At 18 somites (C) and 24 hpf (D), *pax3a* expression is observed within the midbrain, hindbrain and dorsal progenitors (dP) within the spinal cord. Transcription is decreased in the posterior neural tube at 32 hpf (E) and absent at 48 hpf (F), however expression is maintained anteriorly.

4.2 – Pax7 paralogs are expressed upon neural tube formation

Pax7a transcription was first observed in the presumptive midbrain of the zebrafish embryo at the 6 somites stage (Fig. 4.2A) and this restricted pattern of mRNA localisation remained evident until the 12 somites stage (Fig. 4.2B). Expression was markedly increased within the midbrain and hindbrain at the 18 somites stage and mRNA was detected within anterior regions of the spinal cord following neurulation (Fig. 4.2C). Both wholemount preparations and transverse sections of trunk tissue at 24 hpf exhibited *pax7a* transcription within the neural progenitors of the zebrafish dorsal spinal cord (Figs. 4.2D and 4.3B). However mRNA was observed to be excluded from the most dorsal progenitor cells in cross sections of the spinal cord (Fig. 4.3B). *Pax7a* mRNA was absent from neural progenitors in the tail bud region of the posterior spinal cord at 24 hpf (Fig. 4.2D). Expression was decreased in the anterior spinal cord at 32 hpf and varying levels of transcript were observed in adjacent cells (Fig. 4.2E). High levels of *pax7a* mRNA were observed in cells within the posterior spinal cord at 32 hpf, in contrast to data obtained at 24 hpf. Expression was also observed within migratory neural crest cells in the somites and the eye. At 48 hpf, *pax7a* was transcribed in the midbrain and hindbrain but mRNA was not detected in the spinal cord (Fig. 4.2F).

Pax7b expression in the early embryo was restricted to the developing midbrain until the 12 somites stage, in a pattern that closely correlated to that observed for *pax7a* (Figs. 4.4A and B). Transcriptional induction of *pax7b* following neurulation was comparable to that observed for *pax7a*, as expression was evident in the midbrain, hindbrain and most anterior regions of the spinal cord at the 18 somites stage (Fig. 4.4C). By 24 hpf, transcription was detected in the dorsal domains of the midbrain, hindbrain and spinal cord but was excluded from the tail bud (Fig. 4.4D). *Pax7b* expression was maintained at 32 hpf in midbrain, hindbrain and a small population of cells in the posterior region of the spinal cord (Fig 4.4E). No transcription was detected in embryos assessed at 48 hpf (Fig 4.3F).

The expression profile of Pax7 paralogs during zebrafish spinal cord patterning closely resembles that observed during mouse embryogenesis (Jostes et al., 1991). Accordingly, transcription is first observed following the completion of neurulation at trunk level and is excluded from the most dorsal progenitors within the spinal cord. This profile is in contrast to Pax7 expression during chick CNS development, which shares similar temporal dynamics but span all dorsal progenitors (Kawakami et al., 1997).

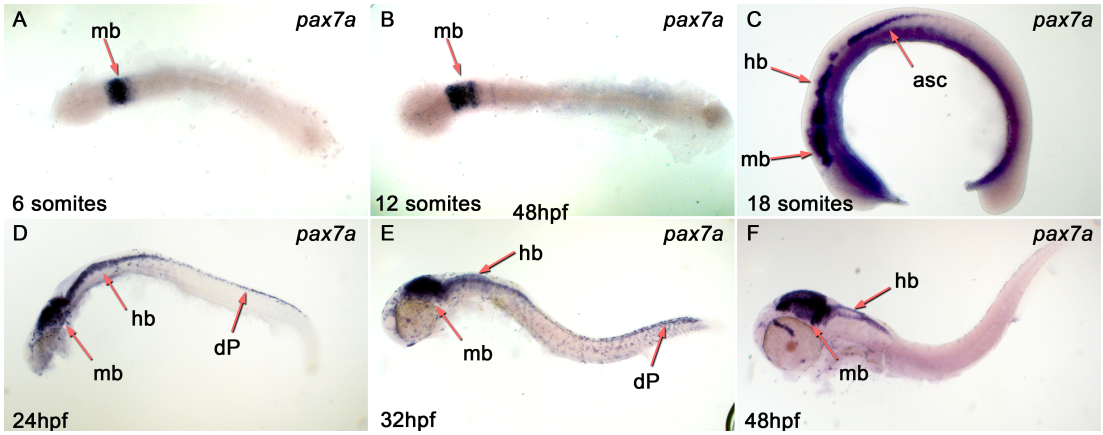


Fig. 4.2 – The profile of *pax7a* expression during zebrafish CNS development

Pax7a transcription is restricted to the presumptive midbrain (mb) at both the 6 (A) and 12 somites stages (B). Following neurulation, transcripts are detected in the midbrain, hindbrain (hb) and anterior spinal cord (asc) (C). At 24 hpf, *pax7a* is expressed throughout the midbrain, hindbrain and dorsal progenitors (dP) within the neural tube but is excluded from the tail bud (D). Transcription is reduced in the anterior spinal cord at 32 hpf but induced posteriorly within the tail bud (E). At 48 hpf, *pax7a* is not expressed in the spinal cord but is maintained within the midbrain and hindbrain (F).

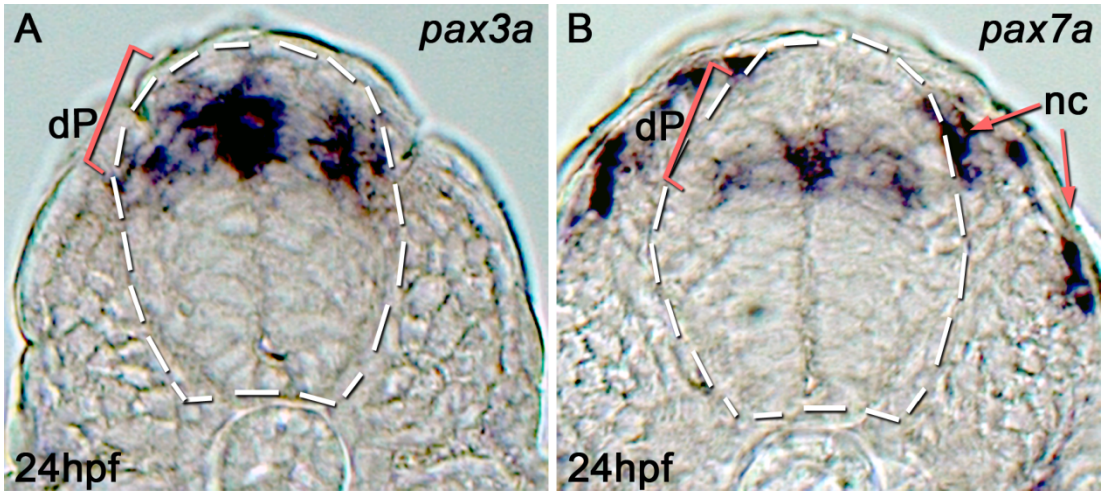


Fig. 4.3 – The distribution of *pax3a* and *pax7a* transcripts within the spinal cord

Transverse sections of the zebrafish neural tube at trunk level reveal *pax3a* transcription in all dorsal progenitors (dP) within the neural tube (A). By contrast, *pax7a* transcripts are excluded from the roof plate and the most dorsal progenitor populations (B). In addition, *pax7a* transcription is detected in neural crest (nc) cells that have delaminated from the dorsal neural tube (B).

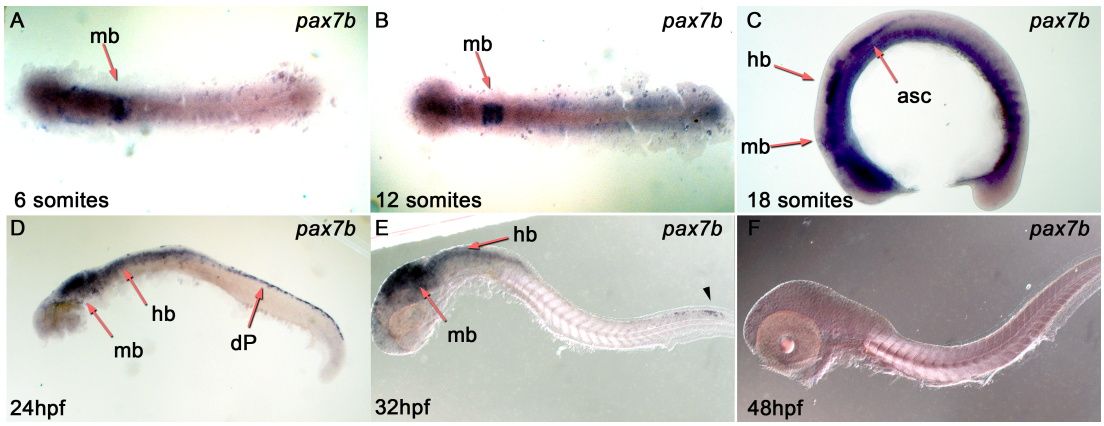


Fig. 4.4 – The profile of *pax7b* expression during zebrafish CNS development

Pax7b transcription is restricted to the presumptive midbrain at 6 (A) and 12 somites stages (B). Expression is restricted to the midbrain, hindbrain and anterior neural tube at 18 somites (C). At 24 hpf, *pax7b* is expressed in the dorsal midbrain, hindbrain and neural tube but is excluded from the most posterior region of the tissue (D). Transcripts are absent from the anterior neural tube at 32 hpf but are detected at the posterior limit of the CNS (arrowhead in E). *Pax7b* is not expressed at 48 hpf (F).

4.3 – Assessing Pax3/7 protein localisation in zebrafish embryos

A potential limitation of using zebrafish embryos in a study of this type is the lack of antibodies availiable for use in immunohistochemistry, as only a subset of those validated for use on mouse and chick tissue recognise the zebrafish antigens. Previous investigations have demonstrated that a monoclonal Pax7 antibody can be used to visualise protein localisation in zebrafish embryos (Patterson et al., 2010, Stellabotte et al., 2007), however there are currently no antibodies raised specifically against Pax3 that are suitable for immunostainings. In order to assess the distribution of both Pax3 and Pax7 proteins during CNS development, we sought to utilise two monoclonal antibodies, DP311 and DP312, raised against the highly conserved homeodomain of the *Drosophila* Pax3/7 class gene *paired* (Davis et al., 2005). Previous reports have indicated that DP312 has a far greater specificity for the recognition of Pax3/7 class genes, due to the ability of DP311 to tolerate amino acid substitutions in position 29 of the antigen homeodomain (Davis et al., 2005). Consequently, DP312 has been used as a marker of Pax3/7 protein localisation in studies of myogenesis (Feng et al., 2006; Hammond et al., 2007; Koudijs et al., 2008; Seger et al., 2011) and neural crest specification (Curran et al., 2009; Minchin et al; 2008) in zebrafish embryos. Recognition of the zebrafish Pax7 protein by DP312 has been biochemically verified by comparison of western blots obtained from whole embryo extracts probed with either DP312 or Pax7, revealing a 50kDa Pax7 complex bound by both antibodies (Hammond et al., 2007). Pax3 binding was demonstrated by the presence of a 55kDa Pax3 complex in embryo extracts probed with DP312 and the subsequent loss of this interaction in Pax3 morphants (Minchin et al., 2008). Although DP312 is frequently used as a marker of Pax3/7 protein localisation in zebrafish embryos, it has not been employed in studies of CNS patterning. Consequently, we sought to examine the specificity of both DP311 and DP312 in cross sections of zebrafish trunk tissue to determine their suitability for use in this investigation.

DP311 immunostaining strongly labelled subpopulations of superficial and medial slow muscle precursor cells within the somites, consistent with the described role of Pax3/7 in these cells (Fig 4.5A). Within the CNS, strong immunoreactivity was observed in the Pax3/7 expression domain of the dorsal spinal cord as expected. Localised DP311 staining was also detected in the FP at the ventral midline of the spinal cord. This ectopic labelling was most probably due to antibody cross-reactivity with Arx, a well-defined marker of the FP that encodes a *paired* type homeodomain

(Miura et al., 1997). In support of this conclusion, DP311 has previously been shown to recognise *aristaless*, the *Drosophila* homolog of vertebrate Arx (Davis et al., 2005). Comparatively weaker staining was observed in the intermediate region of the spinal cord in a pattern consistent with the domain of Pax6 expression. Similar patterns of DP311 immunostaining were produced in spinal cord sections of chick and mouse embryos assessed at comparable stages (Figs. 4.5B and C).

DP312 staining highlighted a subset of superficial slow muscle cells in the somite but was not observed in the medial pioneer population of muscle precursors (Fig 4.6A). Immunoreactivity was detected within the dorsal Pax3/7 domain along the entire anterior-posterior axis of the CNS. Single cells within the presumptive pMN domain of the ventral spinal cord were labelled by DP312 in more anterior sections of spinal cord, possibly due to a cross reaction with Pax4 (Nikolaos Balaskas, personal communication ; Sosa-Pineda et al., 1997). This ventral domain of DP312 reactivity was greatly expanded within the hindbrain and included all cells within the presumptive p3 and pMN domains, however it did not extend to the floor plate (Fig. 4.6B). Staining was also observed within the Pax3/7 domain of the dorsal hindbrain and the two domains were clearly delineated. Similar profiles of DP312 immunoreactivity were observed in transverse sections of chick and mouse spinal cord tissue assessed at comparable stages (Figs. 4.6C and D).

Taken together, these results indicate that DP311 is too broadly cross-reactive within the CNS to be employed in this study as a marker of Pax3/7 protein. The observed cross reactivity with Pax6 is particularly problematic as its expression domain overlaps with that of Pax3/7, making a clear delineation of the Pax3/7 ventral boundary impossible. In contrast, DP312 staining appears to reproducibly label the dorsal half of the CNS and shows no cross reactivity with Pax6 in the three sources of tissue examined. We reasoned that the single cells labelled in the ventral half of the anterior spinal cord are easily discriminated from the Pax3/7 domain and concluded that DP312 staining provided sufficient specificity to be used as a marker of Pax3/7 protein in this investigation.

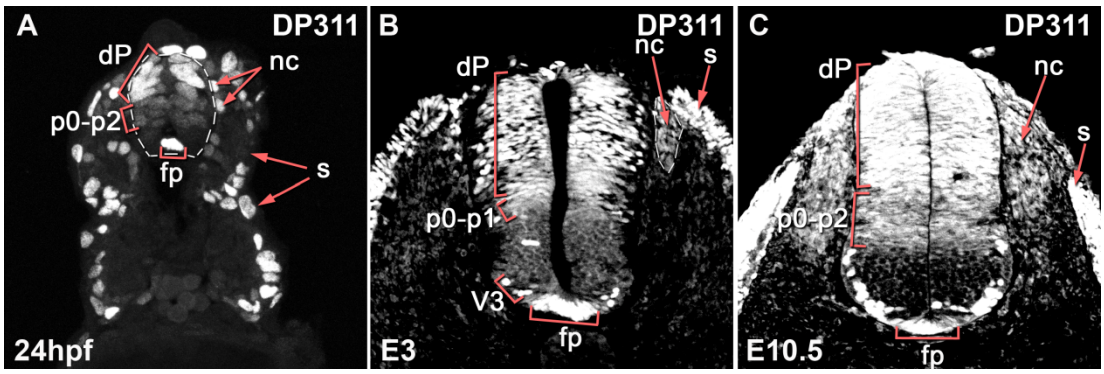


Fig. 4.5 – DP311 is broadly cross-reactive in the neural tube

DP311 immunoreactivity is observed in superficial and medial populations of slow muscle progenitors in the somite (s). Within the neural tube, the Pax3/7 domain of dorsal progenitors (dP), the Pax6 domain (p0-p2) and the floor plate (fp) are labelled by DP311 (A). Staining is also observed in delaminating neural crest cells (nc) (A). DP311 staining labels the Pax3/7 domain, the floor plate, populations of ventral interneuron progenitors and ventral interneurons (V3) within the chick neural tube (B). In mouse tissue, DP311 labels the Pax3/7 and Pax6 expression domains , the floor plate and subpopulations of motor neurons (C).

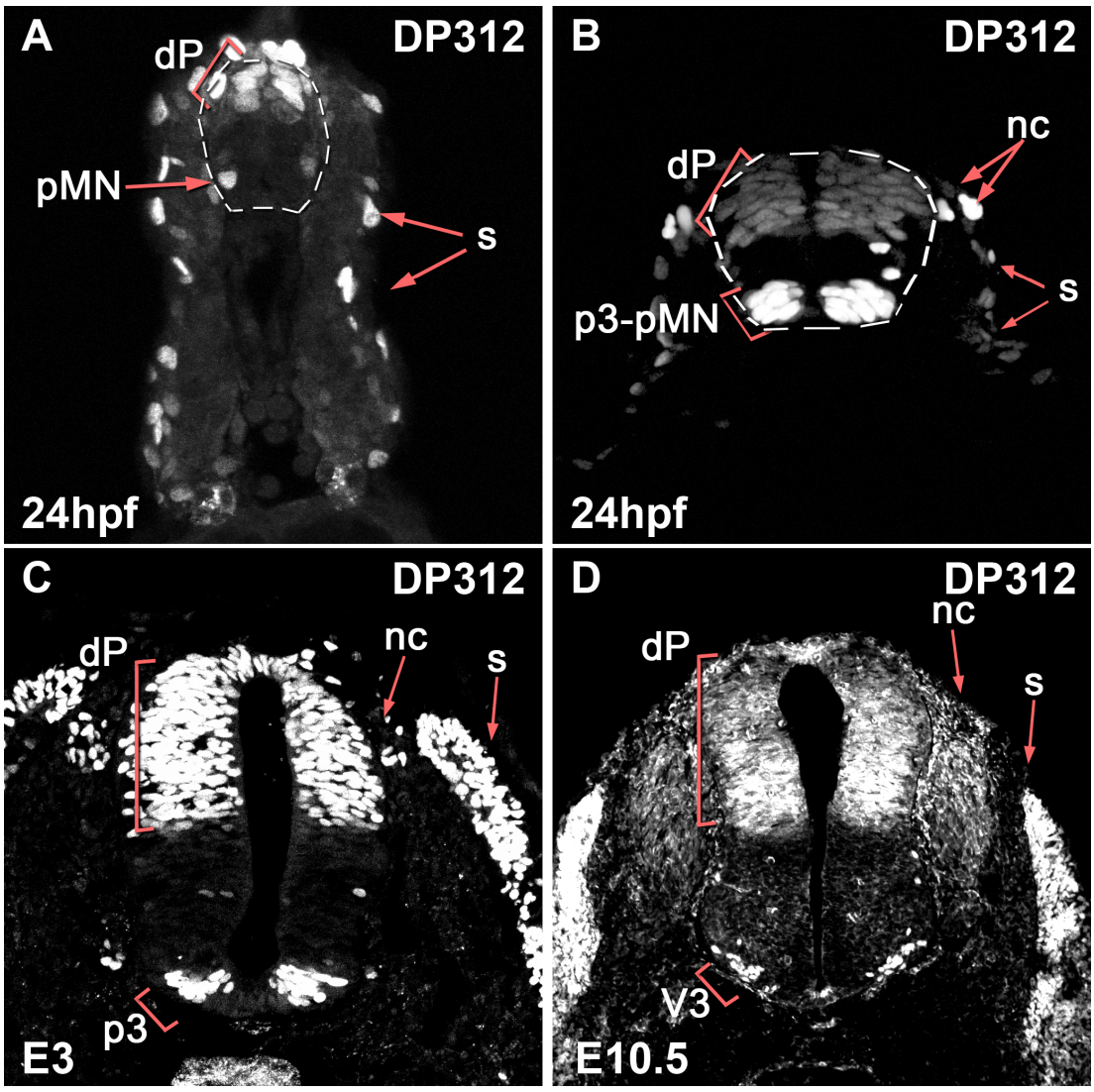


Fig. 4.6 – DP312 demonstrates specificity for Pax3/7 recognition *in-vivo*

At trunk level, DP312 labels slow muscle fibes within the somites (s), the Pax3/7 domain (dP) of the neural tube and single cells within the pMN (A). Within the midbrain, DP312 marks all cells within the pMN and p3 domains (p3-pMN) in addition to the Pax3/7 domain (dP) and neural crest (nc) (B). In chick (C) and mouse tissue (D), DP312 staining labels the Pax3/7 domain and both progenitors (p3) and neurons (V3) ventrally.

4.4 – Characterisation of Pax3 and Pax7 BAC transgenic lines

In order to gain insight into the molecular regulation of Pax3 and Pax7 we established a collaboration with Professor Philip Ingham, whose laboratory has previously studied the roles of these genes during muscle development. Among the tools created during these studies were bacterial artificial chromosome (BAC) transgenic reporter zebrafish lines that have been shown to recapitulate the expression of *pax3a* and *pax7a* during myogenesis and muscle regeneration (Seger et al., 2011). This was achieved by a two-step recombination strategy that first inserted an EGFP reporter into the transcription start site of the selected locus. Subsequently, a kanamycin resistance gene flanked by IsceI meganuclease sites was targeted to the chloramphenicol resistance gene present within the BAC. Previous studies performed in zebrafish have demonstrated that the addition of meganuclease sites greatly improves the germline integration of large constructs, such as BACs (Grabher et al., 2004 ; Thermes et al., 2002). Studies in the Ingham laboratory have demonstrated that these targeted BACs contain regulatory elements that are sufficient to drive transgene expression within the developing dermomyotome and satellite cells of mature myofibres (Seger et al., 2011). Transgene expression within CNS tissues had previously been observed during the establishment of these lines, however a detailed analysis of their expression within the spinal cord has not been performed.

DKEY-20F20 is a 247 kb fully sequenced BAC, Genbank accession number BX085193, that contains the entire zebrafish *pax3a* coding region. A second locus contained within the BAC encodes Acetyl-CoA synthetase family member 3B, an enzyme with no defined expression pattern. Analysis of GFP immunostained wholemount preparations of *Pax3 - GFP*^{l150} embryos at 24 hpf revealed a clear correlation between the activity of the transgene and the previously defined distribution of *pax3a* transcript (Fig. 4.7A). GFP labelling was observed within the somites of transgenic embryos and throughout the entire anterior-posterior axis of the CNS, including the forebrain, which does not appear to express *pax3a* mRNA at this stage. Transverse sections revealed that the activity of the transgene was dorsally restricted in neural progenitors and a population of commissural interneurons within the spinal cord. Antibody stainings with DP312 demonstrated that the cellular localisation of the BAC driven GFP transgene and Pax3/7 completely overlapped in both neuronal progenitors and presumptive neural crest cells (Fig. 4.7B). These results indicate that this transgenic line recapitulates the endogenous expression profile of *pax3a* within the spinal cord and

that DP312 immunostaining is highly specific for this domain. However, the ectopic expression of GFP within the forebrain suggests that additional CRMs are required to silence *pax3a* expression in this tissue.

Pax7-GFP^{I131} stable transgenic lines were created by modification of the CH73-62K19 BAC. Alignment of the sequenced BAC ends suggested that the entire Pax7 locus, 45 kb of upstream and 105 kb of downstream sequence are contained in this genomic interval. Analysis of this stable line at 24h demonstrated that the expression of *pax7a* mRNA in anterior regions of the CNS was recapitulated in all embryos assessed (Fig. 4.7C). However, transgene activity within the spinal cord was not visible in live embryos or wholemount preparations stained for GFP protein. Sections of spinal cord tissue revealed that this BAC was sufficient to drive expression in the dorsal neural tube with a ventral limit that coincided with that of DP312 staining (Fig. 4.7D). The most dorsal regions of the spinal cord was not labelled by GFP staining in this stable line, as demonstrated by the presence of a population of cells that were solely DP312 positive. These data are consistent with *in-situ* hybridisation results, which demonstrated that *pax7a* transcription is excluded from the most dorsal progenitor cells at spinal cord level (Fig. 4.3B).

Taken together, these data indicate that the cis-regulatory modules sufficient for *pax3a* expression in the spinal cord are present in the genomic interval contained in the BAC DKEY-20F20. Similarly, CH73-62K19 BAC transgenics recapitulate *pax7a* mRNA expression in spinal cord progenitors. The varied levels of transgene expression observed in the CNS of each stable line might be due to the integration site of the BAC into the genome.

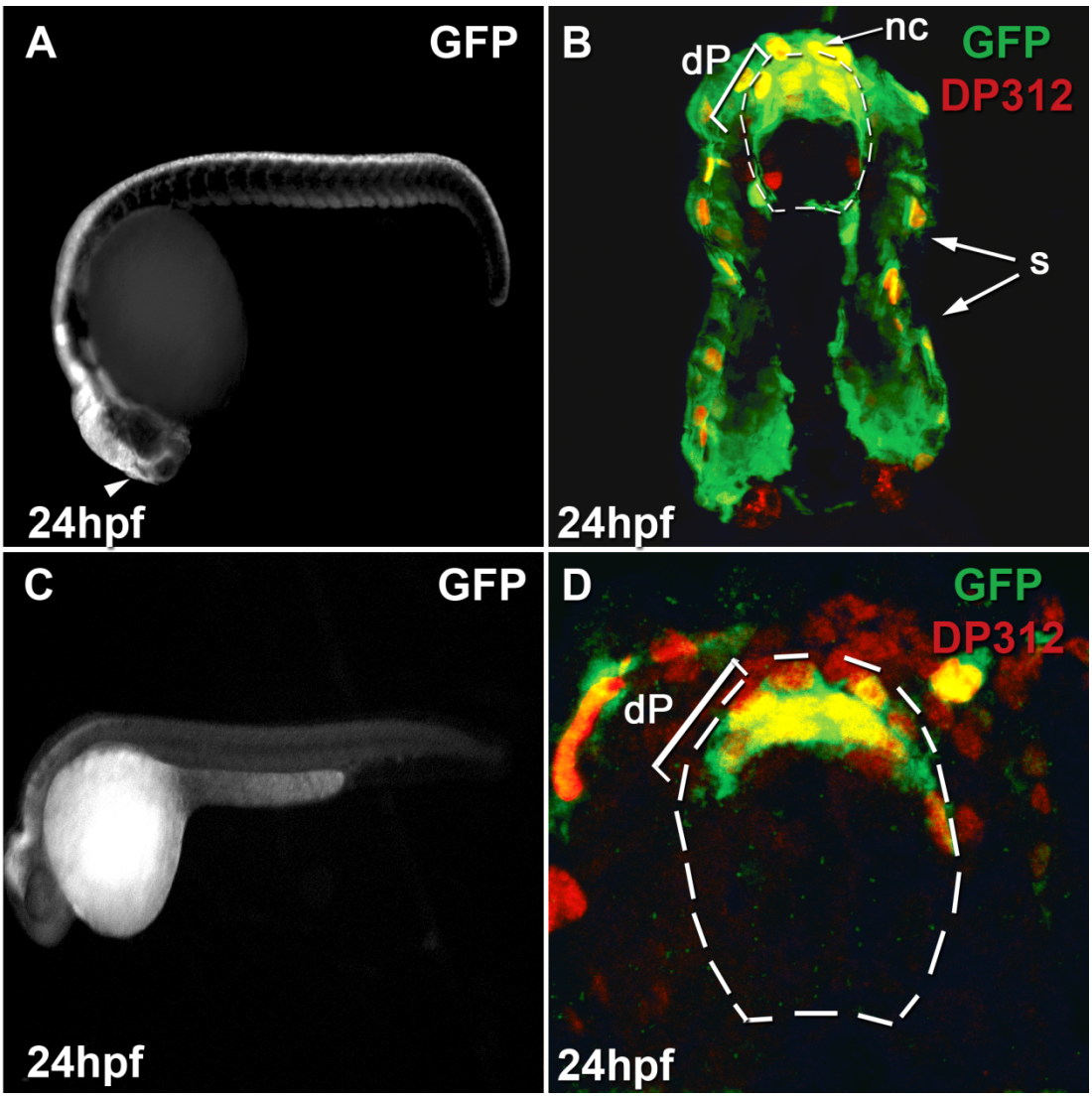


Fig. 4.7 – Zebrafish BAC transgenic lines recapitulate Pax3/7 expression in the spinal cord

The Pax3-GFP^{I150} stable line recapitulates the expression profile of *pax3a* in the spinal cord at 24 hpf (A). However, ectopic GFP staining is observed within the developing forebrain (arrowhead in A). Transverse sections at trunk level reveal that GFP co-localises with DP312 staining in dorsal progenitors (dP) within the spinal cord, presumptive neural crest (nc) and the somites (s) (B). The Pax7-GFP^{131} line displays transgene labelling within the midbrain and hindbrain, however GFP is not observed within the spinal cord in whole mount preparations (C). A population of Pax7-GFP positive cells co-localise with DP312 staining in transverse sections of the neural tube, however the most dorsal Pax3/7 expressing cells are GFP negative (D). These data are consistent with the defined expression profiles of *pax3a* and *pax7a* in the spinal cord.

Chapter 5: Identifying and testing putative enhancers

5.1 – Comparative genomic analyses of the Pax3 and Pax7 loci

In order to investigate the molecular mechanisms that determine the expression of Pax3 and Pax7 during CNS development, it was first necessary to identify putative regulatory regions associated with each locus. Several methods of CRM identification are currently available, discussed in Chapter 1.2. These techniques vary greatly in term of the scale of experiment performed and the amount of data produced. As a result, studies examining the molecular basis of gene regulation can range from the dissection of a known regulatory module to genome wide profiling of chromatin states in multiple cell types or species.

Phylogenetic conservation of non-coding sequence was chosen as our preferred method of CRM discovery as it provides a relatively rapid means of identifying regions of interest associated with a single locus. The application of comparative genomics for the purpose of CRM discovery has been shown to be particularly suited to studies investigating developmental genes, such as Pax3 and Pax7 (Woolfe and Elgar, 2008). This observation is hypothesised to be due to the highly conserved expression profile of these transcription factors across the vertebrate subphylum during embryogenesis. We therefore reasoned that the *in-silico* identification of CNEs, followed by *in-vivo* assessment of their regulatory potential would be an appropriate means of defining a set of CRMs for molecular dissection.

Several publically available genome browsers host comparative genomic alignments, with the aim of providing a source of putative regulatory elements for studies such as ours (Engstrom et al., 2008 ; Ovcharenko et al., 2004). These databases were created by pairwise alignments of whole genome assemblies and currently do not host any data concerning the functions assigned to these regions. In addition to these genome wide resources, several individual research groups curate CNE browsers that provide comparative genomic alignments focused on a selective set of developmental loci (Visel et al., 2007 ; Woolfe et al., 2007). Although these databases are derived from varied data sources and utilise different methods of sequence alignment, *in-vivo* evidence of CRM activity in the groups' preferred model organism is provided, when available. Alignments of both the Pax3 and Pax7 loci are available in each of the described databases, however none of the CNEs identified had been functionally annotated at the time of this investigation. We sought to perform our own stringent

comparative alignments of the Pax3 and Pax7 loci, in order to create a set of CNEs for further analysis. We reasoned that the wealth of publically available genome alignments would allow us to validate our dataset before proceeding with functional analysis *in-vivo*.

The large evolutionary distance typically employed in comparative genomics has been shown to enhance the predictive power of this technique. However it also poses difficulties in ensuring that all CNEs associated with a locus are identified, due to genome inversions and local shuffling events that occur during evolution (Ovcharenko et al., 2005). We chose to utilise the Mulan suite of comparative genomics tools as our preferred method of CNE identification because it takes these issues into account by performing a two-step alignment strategy. First, the evolutionary relationship between the input dataset is determined by calculating the number of substitutions per kb of sequence by pairwise BLAST and represented as an editable phylogentic tree. The description of how the input sequences are related to each other is then used to guide the threaded blockset alignment algorithm, which locally aligns regions of homology between the subject and query genomes independent of their position or orientation within the input data (Ovcharenko et al., 2005). This implementation of comparative genomics is therefore optimised for the identification of CNEs associated with a single locus over large evolutionary distances.

Our initial strategy for the selection of genomic regions for analysis was based upon the observation that the zebrafish Pax3 and Pax7 paralogs are expressed in overlapping domains within the spinal cord. We therefore hypothesised that these loci might share common CRMs with sufficient primary sequence homology to be identified by comparative genomics. We discovered that that was not the case, as no CNEs could be identified in pilot experiments that attempted to align these four loci from the zebrafish genome (data not shown). We therefore attempted to align the human genomic sequence (Hg18), spanning the desired locus and the surrounding intergenic regions, to the corresponding regions of the mouse (Mm9) and either the duplicated Pax3 or Pax7 loci in the zebrafish (Zv7) genome. We discovered that it was possible to identify CNEs within the Pax7 locus employing this strategy and continued to use this alignment method in our subsequent analysis. By contrast, we observed that identification of CNEs associated with Pax3 was not possible when sequence containing the zebrafish *pax3b* locus was included in the source dataset (data not shown). In order to avoid this issue, yet retain the stringency we desired in our genomic analysis, we chose to replace the zebrafish *pax3b* locus with sequence representing the *pax3a* locus in the Fugu (Fr2)

genome. We found that CNEs were readily identified in these alignments and chose to pursue this strategy for further analysis of the Pax3 locus.

Processed data was viewed in the Mulan visualisation tool, permitting a graphical representation of the location of CNEs and their degree of conservation. The parameters imposed on the data at this point to improve the stringency of our alignments are described in Chapter 2.2. Conserved TFBS present in each CNE were annotated using the MultiTF tool of the Mulan suite (Loots and Ovcharenko, 2007). Any CNEs that did not contain conserved TFBS were reasoned to be repetitive sequence and discarded in this process. This analysis revealed the presence of 6 CNEs in the large fourth intron of Pax3, ranging in size from 150 to 500 bp (Fig. 5.1). Similarly, 6 CNEs between 50 and 450 bp were associated with the Pax7 locus. These regions were located both proximally to the transcription start site and within the third intron of the gene (Fig. 5.2).

In order to validate our genomic analysis, we interrogated the Condor database for the presence of identified CNEs by BLAST (Woolfe et al., 2007). This comparison revealed that all CNEs in our dataset overlapped with regions identified by the Condor database. This result was expected, as our alignment strategy was based upon the parameters previously employed during the creation of the Condor dataset. We often found that the larger CNEs identified by our Mulan alignments were represented as clusters of several individual Condor elements, separated by small intervals of genomic sequence. One such example is Pax3-CNE2, which when compared with the Condor data shows homology to two CNEs that are separated by 38 bp of sequence that is not recognised as conserved. Accordingly, Positions 20 – 148 of CNE2 are homologous to CRCNE00001996, whereas positions 186 – 263 are homologous to CRCNE00001995. We reasoned that the discrepancies between our dataset and the Condor database could be produced by a combination of the source of input data and the method of alignment. Our study searched for CNEs shared between two mammalian genomes and two fish genomes, whereas the Condor dataset is based upon the alignment of three mammalian genomes to the fugu sequence. The difference in input data coupled with the choice of algorithms employed to process the alignment may explain why although all of our CNEs are present in the Condor dataset, we identify fewer individual CNEs. The co-ordinates of the regions selected for functional analysis in the current release of the zebrafish genome (Zv9) and the corresponding CNEs in the Condor database are listed in Table 5.1.

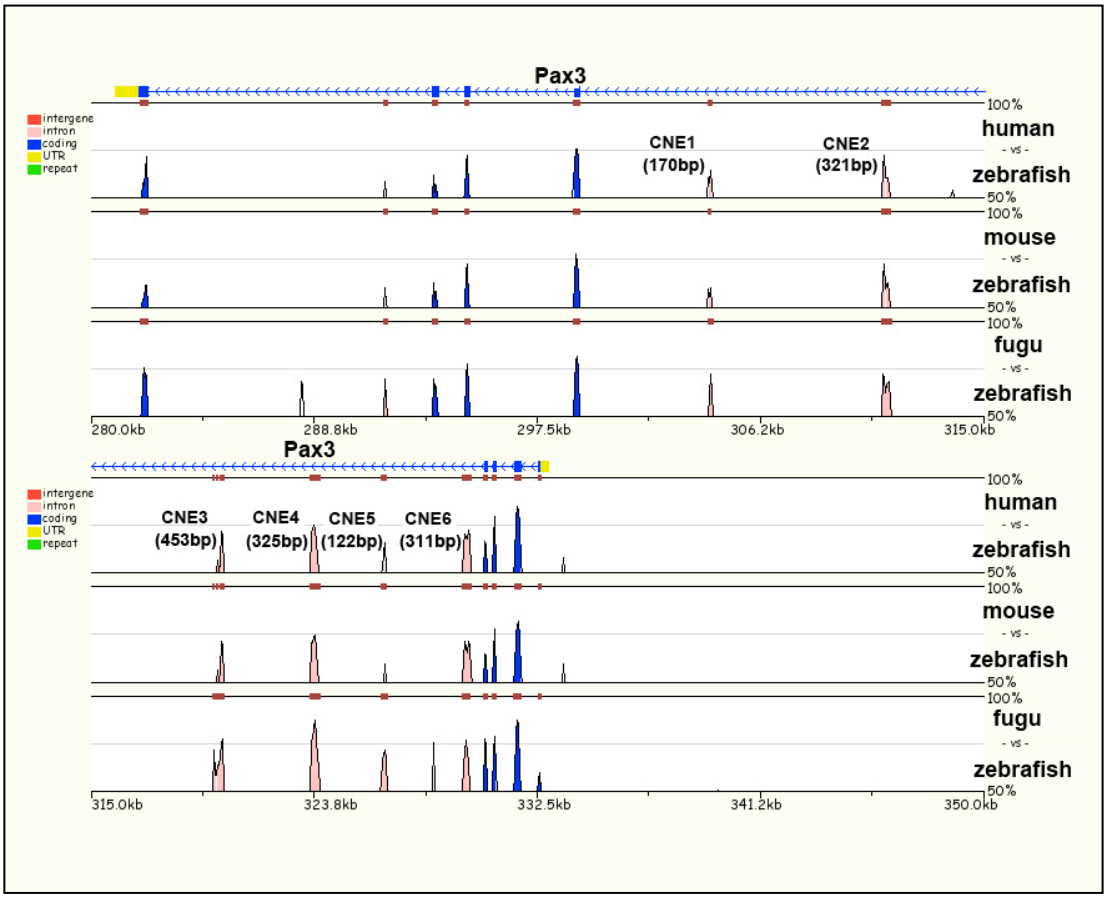


Fig. 5.1 – Comparative genomic analysis of the Pax3 locus

Mulan alignment of sequences representing the Pax3 locus in the human (hg18), mouse (mm9), zebrafish (Zv7) and fugu (Fr2) genomes. Exons are labelled in blue and CNEs are pink. The 6 CNEs identified for functional assessment are labelled within the plot.

Figure cropped from original alignment for clarity.

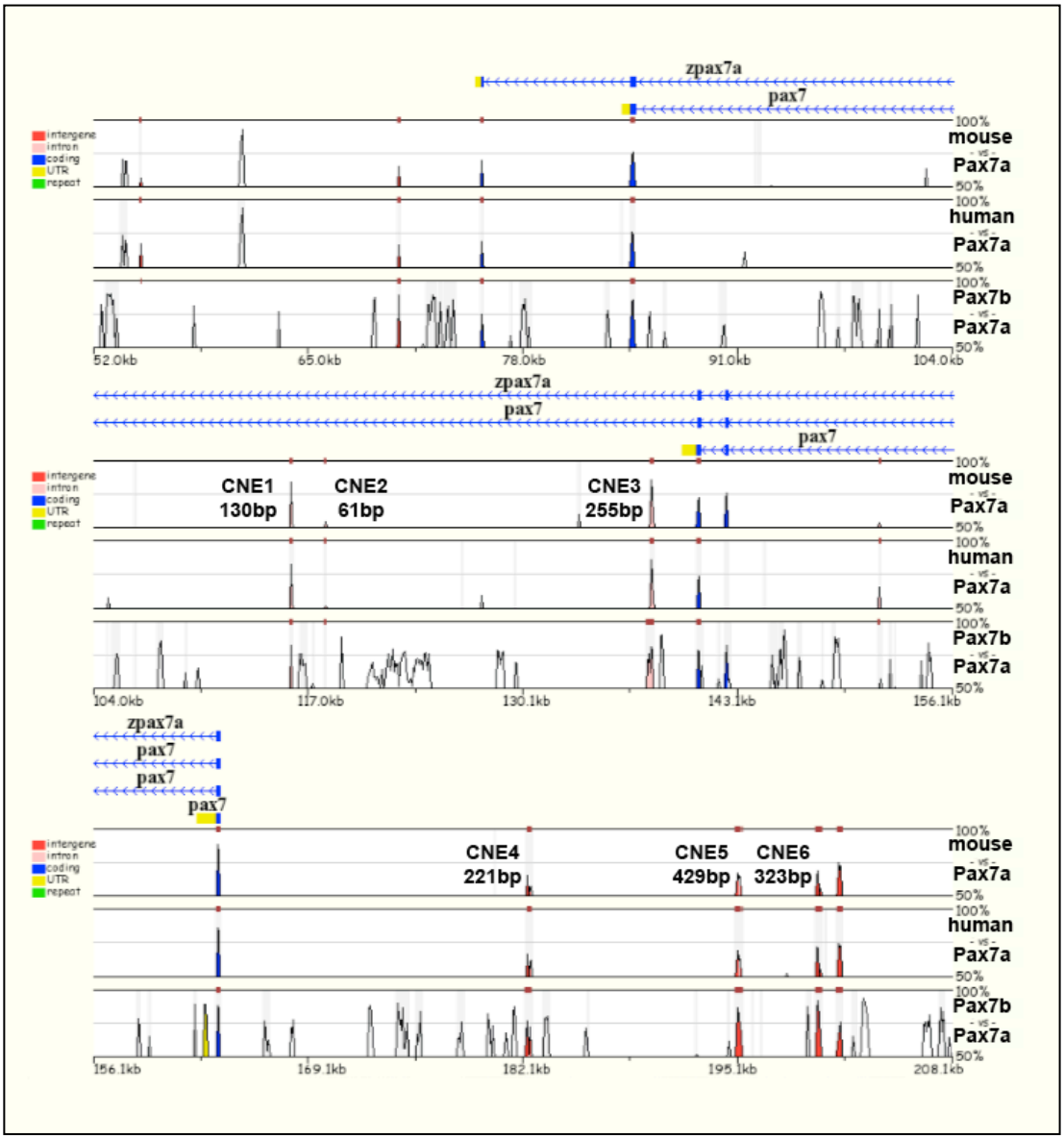


Fig. 5.2 – Comparative genomic analysis of the Pax7 locus

Mulan alignment of the human (hg18) and mouse (mm9) Pax7 loci with zebrafish *pax7a* and *pax7b* (Zv7). Exons are labelled in blue, intergenic CNEs are red and intragenic CNEs are pink. The 6 CNEs identified for functional assessment are labelled within the plot.

Figure cropped from original alignment for clarity.

Identifier	Genomic Location (Zv9)	Overlapping Condor CNEs
Pax3-CNE1	Ch2: 47609549-47609886	CRCNE00001997
Pax3-CNE2	Ch2: 47602558-47603078	CRCNE00001995 and CRCNE00001996
Pax3-CNE3	Ch2: 47593693-47594345	CRCNE00001990 and CRCNE00001994
Pax3-CNE4	Ch2: 47589982-47590506	CRCNE00001989
Pax3-CNE5	Ch2: 47587333-47587654	CRCNE00001987
Pax3-CNE6	Ch2: 47583981-47584491	CRCNE00001984 and CRCNE00001985
Pax7-CNE1	Ch11: 42451808-42452137	CRCNE00009579
Pax7-CNE2	Ch11: 42453941-42454201	CRCNE00009578
Pax7-CNE3	Ch11: 42473594-42474048	CRCNE00009577
Pax7-CNE4	Ch11: 42518191-42518611	CRCNE00009574 and CRCNE00009575
Pax7-CNE5	Ch11: 42530820-42531448	CRCNE00009572, CRCNE00009573 and CRCNE00009587
Pax7-CNE6	Ch11: 42535678-42536200	CRCNE00009570

Table 5.1 – CNE location in the zebrafish genome and equivalent Condor elements

The location of each CNE identified by comparative genomics in the Zv9 release of the zebrafish genome, determined by BLAST, is listed. Many of the CNEs identified by our genomic analysis are represented in the Condor database as several individual elements.

5.2 – Testing CNE activity *in-vivo*

We sought to investigate the regulatory potential of each CNE *in-vivo*, with the aim of reducing the complex spatiotemporal expression patterns of Pax3 and Pax7 to the function of a set of discrete CRMs. Individual CNEs were amplified from freshly prepared BAC DNA and cloned into the MiniTol2-Citrine expression vector (Fig. 5.3) according to the protocol described in Chapter 2.2. This expression vector was selected for use in this investigation as it was designed to facilitate the screening of putative regulatory elements in zebrafish embryos (Javier Terriente, personal communication).

As described in Chapter 1.2, the activity of a putative enhancer is commonly examined by placing these regions immediately upstream of a minimally active promoter and assaying the transcriptional output with reporter genes. In our assay the function of a CNE was assessed by its ability to modulate the activity of a proximal Thymidine Kinase promoter, which showed no specific activity in control experiments (data not shown). Transcriptional activation is amplified by the Gal4-UAS system (Elliott and Brand, 2008) before being visualised as the production of Citrine, a variant of YFP. This amplification step was desirable for the purposes of this investigation as it ensured that weak or transiently active enhancers could induce sufficient levels of fluorophore protein to be visualised in live embryos. Integration of this reporter into the genome was facilitated by the employment of the Tol2 transposon system, which is frequently used in zebrafish studies to improve the rate of transgenesis (Kawakami, 2007). Messenger RNA encoding the Tol2 transposase was co-injected into a 1-cell stage zebrafish embryo with the MiniTol2-Citrine vector. The transposase is translated *in-vivo* and binds to its minimal recognition sites in the plasmid, resulting in the excision of the expression cassette from the donor vector and its integration in the genome.

Of the 12 CNEs identified by our comparative genomics, 7 were successfully cloned into the MiniTol2-Citrine expression vector and sequence verified. Amplification of all 6 CNEs from the Pax3 BAC DKEY-20F20 was achieved, however CNE6 was not sequence verified. Two CNEs within the Pax7 locus, CNE2 and CNE4, were cloned and sequenced. We were not able to clone the remaining Pax7 CNEs from the CH73-62K19 or CH211-119A10 BACs, the latter being the source from which the Pax7 locus was sequenced during the assembly of the zebrafish genome. The activity of each cloned CNE was assessed *in-vivo* according to the protocol described in Chapter 2.4. Our screen resulted in the identification of 5 putative CRMs that are able to

coordinate transgene expression in a cell type specific manner, all of which are associated with the Pax3 locus. Consequently the major focus of this project became the study of these CRMs and the molecular basis of Pax3 regulation during spinal cord development.

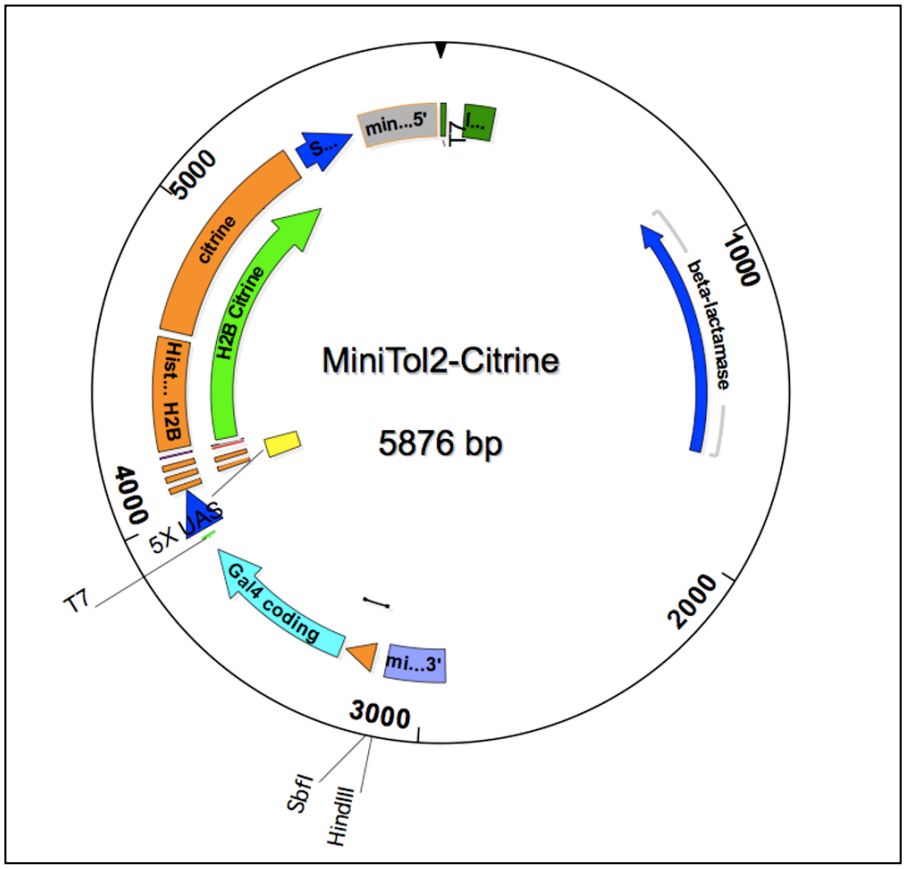


Fig. 5.3 – The MiniTol2-Citrine expression vector

Putative enhancers are cloned into the HindIII-Sbf1 site located proximally to a minimally active Thymidine Kinase promoter. CNE mediated transcription is amplified by the Gal4-UAS system before being visualised by the production of Citrine. The rate of transgenesis is improved by co-injection of Tol2 transposase mRNA, which excises the expression cassette and aids its integration within the genome.

5.3 – CNE2 and CNE4 activate transcription within Pax3 expressing cells

CNE2 is a 321 bp element containing phylogenetically conserved TFBS for multiple bHLH, Homeodomain and retinoic acid response factors (Fig. 5.4). Our *in-vivo* screen suggests that CNE2 may function as an enhancer that is sufficient to activate transcription in specific population of neurons within the spinal cord. At 24 hpf, whole mount preparations of embryos immunostained with a polyclonal anti-GFP antibody revealed Citrine expression specifically within the spinal cord (n=11) (Fig. 5.5A). The profile of reporter gene activation driven by CNE2 was not observed in live embryos, suggesting that it possesses a weak regulatory potential. Higher magnification images of trunk tissue revealed that the cells marked by CNE2 activity appeared to reside within the dorsal half of the spinal cord (Fig. 5.5B). These cells exhibited a polarised morphology and frequently sent processes ventrally towards the floor plate (Fig. 5.5B). The localisation and morphology of cells marked by GFP staining led us to conclude that CNE2 was sufficient to direct transcription within a subset of dorsally located postmitotic neurons. We reasoned that the weak regulatory potential of CNE2 and its activity in postmitotic cells made it an unsuitable candidate for further analysis in this study.

CNE4 is a 325 bp element containing several conserved predicted TFBS for Hox, Oct and Fox transcription factors (Fig. 5.4). In our initial screen at 24 hpf, we observed that CNE4 was sufficient to mediate transcription within a subset of trunk skeletal muscle fibres (n=26) (Fig. 5.6A). We also noted transgene activation in a row of cells located within the dorsal domain of the spinal cord (Fig 5.6A'). The morphology and connectivity of these cells suggested that they might represent Rohon-Beard neurons, a transient population of neural crest derived mechanosensory neurons that are unique to anamniotes (Roberts, 2000). Further molecular characterisation of these cells was not undertaken during this investigation. In anterior regions of the embryo, transgene expression was detected in clusters of cells that were reasoned to be populations of migratory neural crest.

Transverse sections of trunk tissue at prepared at 24 hpf revealed the presence of Citrine expression in isolated populations of medially located fast fibres and superficial slow twitch fibres (n=8) (Fig. 5.6B). Expression was also observed in a number of presumptive neural crest cells located at the dorsal midline of the spinal cord. CNE4 meditated transcriptional activity was present in both the slow and fast twitch classes of

muscle fibres and their nuclei at 48 hpf, as demonstrated in whole mount (Fig. 5.6C) and lateral flat mount preparations of trunk tissue (Fig. 5.6D).

The activity of CNE4 within populations of skeletal muscle cells and neural crest derivatives was encouraging as it suggested that our CNEs represented discrete CRMs whose combinatorial activity defines the expression of Pax3 in multiple tissues. However, we concluded that further analysis of CNE4 was not suitable for this investigation as it did not exhibit activity in spinal cord progenitors. We shared our results with members of the Ingham lab, who created 7 zebrafish transgenic lines that stably expressed CNE4. Citrine expression was observed within both slow and fast populations of muscle fibres in 5 of these lines, however the activity of CNE4 did not appear to be specific and often labelled cells within the developing eye, otic vesicle and anterior CNS (Claudia Seger, personal communication).

Fig. 5.4 – Annotation of conserved TFBS within Pax3-CNEs

MultiTF analysis of CNEs associated with the Pax3 locus reveals that these elements are composed of clusters of phylogenetically conserved TFBS.

Figure cropped from original alignment for clarity.

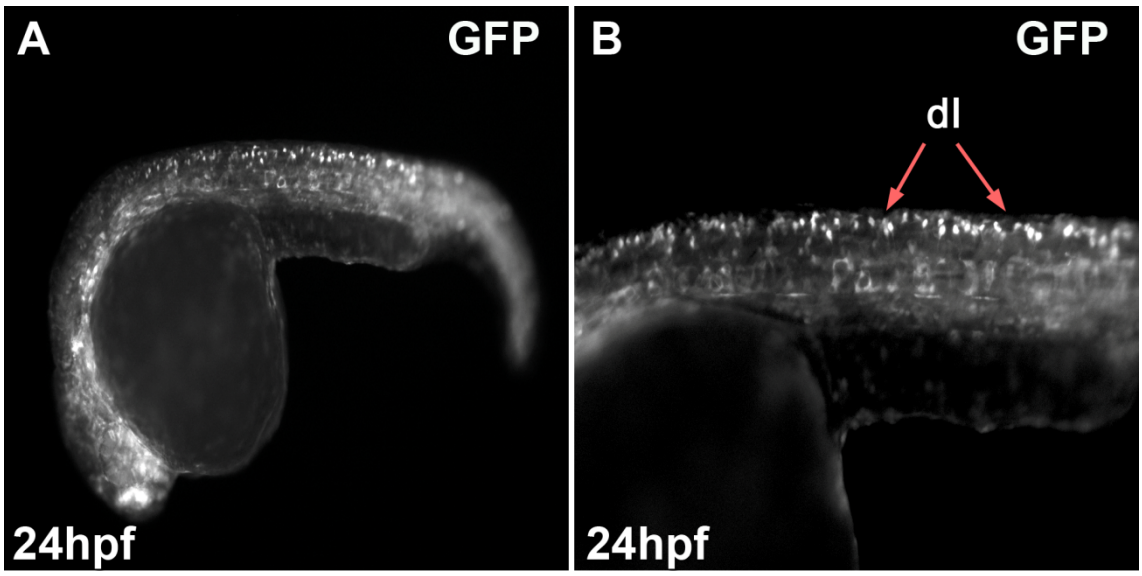


Fig. 5.5 – CNE2 mediated transcription labels dorsal interneurons

Lateral wholemount view of a CNE2 transient transgenic embryo at 24 hpf, stained with a polyclonal GFP antibody. Weak transgene labelling is evident along the AP axis of the dorsal spinal cord (A). These cells are polarised in the DV axis and extend axons ventrally toward the FP, indicating that they represent subpopulations of dorsal interneurons (dI) (B).

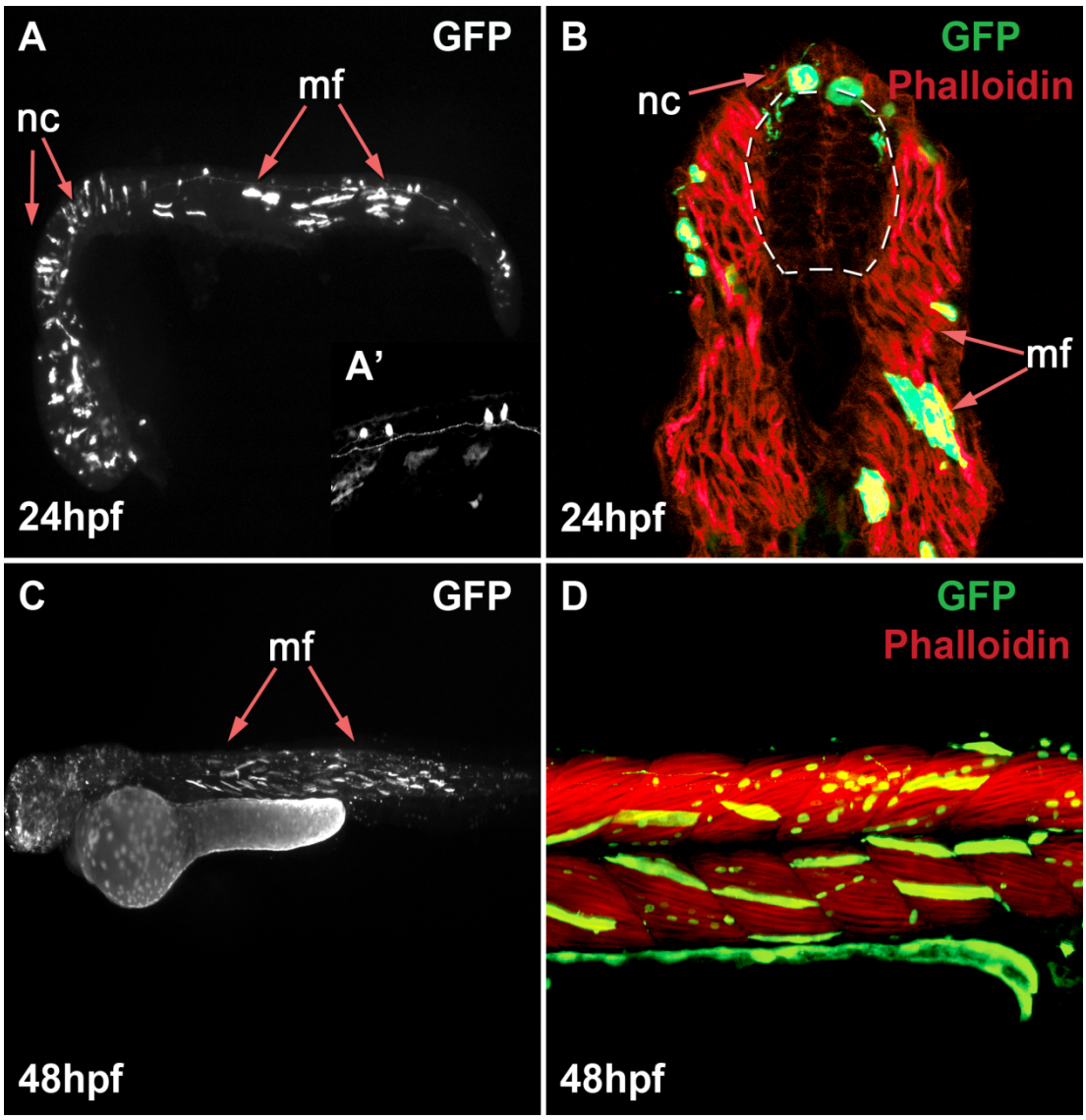


Fig. 5.6 – CNE4 enhances transcription within muscle and neural crest cells

Lateral view of a CNE4 transgenic embryo stained with GFP antibody revealing transgene localisation within muscle fibres (mf), migrating neural crest (nc) (A). Higher magnifation views of trunk tissue reveals GFP labelling of cells with the characteristic morphology of Rohon-Beard neurons (A'). Transverse sections at 24 hpf reveals GFP staining in neural crest cells and muscle fibres within the somite (B). Citrine expression persists within both fast and slow twitch muscle fibres and their nuclei at 48 hpf (C and D).

5.4 – CNE5 is active in the notochord and floor plate

Whole mount preparations of CNE5 transgenic embryos assessed at 24 hpf revealed the extensive GFP labelling of the large columnar cells that comprise the zebrafish notochord, extending from the otic vesicle to the posterior limit of the trunk (n=14) (Fig. 5.7A). We also observed GFP labelling in a row of single cells located dorsally adjacent to the notochord and reasoned that these cells might represent the FP of the neural tube. Mosaic transgene activity was observed in cells within the dorsal spinal cord, hindbrain and forebrain of all embryos assessed at this stage. Transverse sections of trunk tissue at 24 hpf confirmed that CNE5 mediated transcriptional activity labelled cells of the notochord (n=11) (Fig. 5.7B). Citrine was also localised within cells of the FP and a sharp boundary was observed separating these cells from the p3 domain of the spinal cord. Transgene labelling within the somites and neural progenitors was rarely observed and limited to single cells when present. As CNE5 was demonstrated to have such an unexpected regulatory function, we sought to further analyse its activity over the first 72 hours of development in live embryos. We first observed transgene activation within the notochord at approximately the 6 somites stage which persisted until the second day of development (data not shown). The activity of this enhancer appeared to be markedly downregulated by 72 hpf, when only a small number of cells located within the dorsal aorta and cardiac muscle were labelled by GFP antibody staining (n=10) (Fig. 5.7C).

The expression pattern generated by CNE5 transcriptional activity *in-vivo* contrasted with the results obtained from our experiments examining the function of CNE2 and CNE4, as the majority of cells labelled by CNE5 activity do not reside in the Pax3 expression domain. Analysis of the conserved TFBS contained within CNE5 cannot provide a molecular explanation of its activity, as it does not contain sites for key determinants of the notochord or floor plate, such as FoxA2 or Gli family transcription factors (Fig. 5.4). We reasoned that some components of the expression profile modulated by CNE5, such as is observed in the forebrain, hindbrain and cardiac muscle, may be required for the correct profile of Pax3 expression. This hypothesis would require the active silencing of this CRM within the trunk of the embryo during CNS patterning, a mechanism for which there is currently no molecular evidence. Alternatively, this CNE may represent a CRM for a separate gene and function entirely independently of the Pax3 locus.

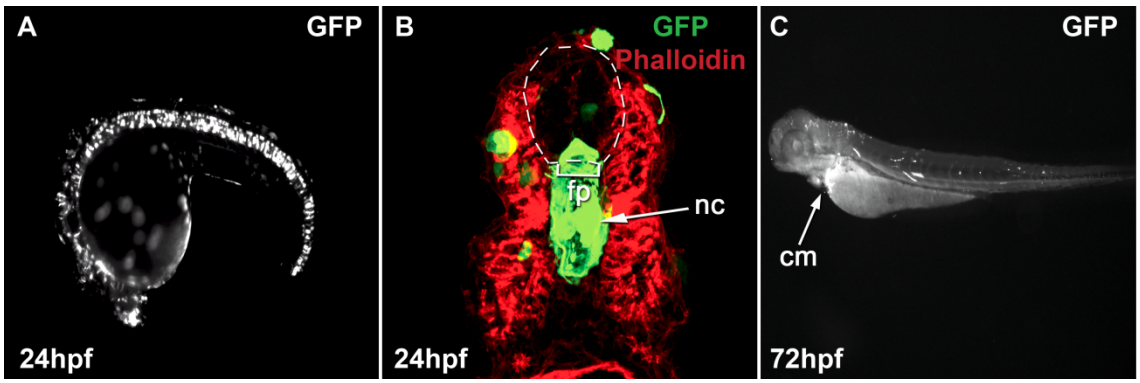


Fig. 5.7 – CNE5 is active in the notochord and floor plate

The large columnar cells of the notochord (nc) are strongly labelled by CNE5 driven transgene expression, assessed at 24 hpf by GFP antibody staining (A). Transverse sections demonstrate that Citrine protein is restricted to the notochord and the floor plate (fp) of the neural tube, with a sharp dorsal boundary at the adjacent p3 domain (B). At 72 hpf, GFP labelling is limited to cardiac muscle (cm) and individual cells within the dorsal aorta (C).

Chapter 6: Characterisation of Pax3 CNE1 activity

6.1 – CNE1 is active within the Pax3 domain of the CNS

CNE1 is a 170 bp conserved element located within the 3' region of Pax3 intron 4 (Fig. 5.1) and is represented in the Condor database as CRCNE00001997 (Table. 5.1). MultiTF annotation of this genomic region suggested that it contained a phylogenetically conserved binding site for the Notch pathway effector, recombination signal binding protein for immunoglobulin kappa J (RBPJ), in addition to sites known to favour the binding of bHLH and homeobox containing transcription factors (Fig. 5.4). Our initial observations of CNE1 mediated transcriptional activity fulfilled the aims of our screen, as we reproducibly observed high levels of transgene production throughout the AP axis of the CNS in live embryos at 24 hpf.

Transgenic samples, fixed at 24 hpf and stained with an anti-GFP antibody, revealed Citrine expression within the hindbrain and midbrain of all embryos assessed (n=67) (Fig 6.1A). Transcriptional activation within the developing forebrain was less reproducible than the robust expression evident within the midbrain or hindbrain and was often very mosaic. Clusters of GFP positive cells were distributed across the entire AP axis of the spinal cord in 98.5% (66/67) of Citrine positive embryos. Higher magnification views of whole mount tissue suggested that CNE1 was transcriptionally active within multiple cell layers of the spinal cord (Fig. 6.1A'). We hypothesised that the majority of these cells resided within the dorsal spinal cord, as they were located adjacent to the dorsal limit of the trunk. We observed two broad classes of cells labelled by CNE1 activity; the first appeared to be polarised in the DV axis and extended ventral processes towards the floor plate. As these cells exhibited a similar morphology to those labelled by CNE2 activity, we concluded that they represented a subset of Pax3 derived interneurons. The second population of Citrine expressing cells did not exhibit any obvious polarisation or processes; we therefore hypothesised that they might be Pax3/7 expressing neural progenitors. Transgene labelling was rarely observed in non-neural tissues and limited to single muscle fibres when present.

Transverse sections of trunk tissue at 24 hpf confirmed that CNE1 mediated transcription was restricted to the dorsal half of the spinal cord in both medially located neural progenitor cells and ventrally projecting neurons (n=10) (Fig. 6.1B). We noted the presence of GFP positive axons extending from the Pax3/7 expression domain that cross the FP to terminate within ventral domains of the contralateral neural tube,

consistent with the expression of Pax3 within the progenitors of commissural interneurons (data not shown). Co-staining of transverse sections with DP312 revealed that the Citrine expression mediated by CNE1 was specifically localised within the Pax3/7 expression domain of the neural tube (n=17) (Fig. 6.1C).

The results of our *in-vivo* analysis of CNE1 activity established this region as an enhancer that possesses sufficient regulatory potential to be readily visualised in the CNS of live transient transgenic embryos, in contrast to the comparatively weak transcription mediated by CNE2. CNE1 driven transcription was demonstrated to be highly specific for the dorsal regions of the zebrafish CNS, in a profile that is consistent with that defined for *pax3a* mRNA and the BAC driven Pax3-GFP reporter described in Chapter 4.4. We were able to demonstrate that the activity of CNE1 within the spinal cord is restricted to the Pax3/7 domain marked by DP312 staining and furthermore, that these cells appear to have a progenitor identity. Taken together, these data suggested that CNE1 was a good candidate for an enhancer that may co-ordinate Pax3 expression in spinal cord progenitor cells. We therefore chose to select this region for further *in-vivo* characterisation, with the aim of determining the spatiotemporal dynamics of CNE1 activity throughout the early stages of CNS patterning.

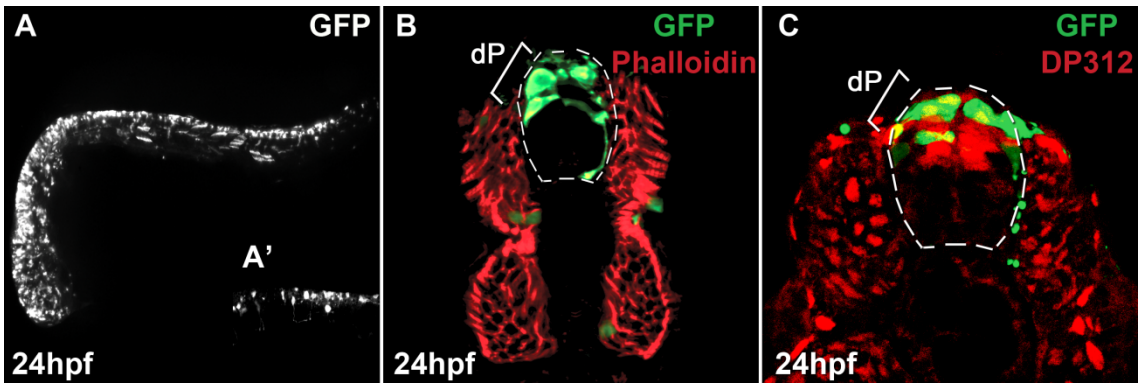


Fig. 6.1 – CNE1 is sufficient to induce transcription within the dorsal neural tube

CNE1 activity is detected along the entire AP axis of the CNS in GFP stained whole mount preparations of embryos assessed at 24 hpf (A). Higher magnification imaging suggests that both progenitors and neurons are labelled by CNE1 activity (A'). Transverse sections of trunk tissue demonstrate that CNE1 activity specifically labels the Pax3/7 domain of dorsal progenitors (dP) within the neural tube (B and C).

6.2 – CNE1 is a CNS specific enhancer

The *in-vivo* activity of CNE1 at 24 hpf suggested that this region may play in a role in the regulation of Pax3 expression, however we were aware that the mosaic transgenesis achieved in our assay imposed limits on our ability to investigate its function. This effect was observed to be particularly problematic during the early stages of CNS patterning, when *pax3a* transcript is first detected and becomes restricted to the lateral neural plate. A series of pilot experiments was able to determine that CNE1 is transcriptionally active at bud stage, however the mosiac nature of the assay made it difficult to draw conclusions on the location of these cells within the early embryo (Data not shown). Given the results of our lineage studies, described in Chapter 3, we reasoned that this stage of development might prove to be critical in the establishment of the Pax3 expression domain. We therefore sought to create zebrafish lines in which the CNE1 expression cassette was stably integrated into the germline, facilitating analysis of this enhancer at early stages of CNS patterning.

A clutch of CNE1 transient transgenic embryos were sorted at 24 hpf on the basis of strong Citrine expression within the spinal cord and maintained at 28.5 °C until the fifth day of development. 30 transgenic embryos were transferred to the care of the aquatics facility, all of which survived to adulthood. Germline integration was assayed in outcrosses of injected founder transgenics to stock wild type strains. The resulting transgenic embryos were sorted from siblings at 24 hpf on the basis of Citrine expression and transferred to the aquatics facility to be maintained as the F1 generation of the stable line. The F0 founder transgenics were maintained in individual tanks to enable a preliminary characterisation of each stable line in outcrosses with wild type strains.

Of the 30 injected fry selected for the creation of stable lines, 3 demonstrated germline integration of the CNE1 expression construct. Citrine expression was readily observed within the CNS tissues of live F2 progeny of each of these lines at 24 hpf, however we noted that the levels of flurophore produced was variable. This was particularly evident in live L1751 derived embryos, which demonstrated a comparatively weaker fluorescence than both L1752 and L1860 transgenics. Whole mount GFP staining of outcrossed F2 embryos at 24 hpf revealed the presence of CNE1 mediated transgene activation along the entire AP axis of the spinal cord in each line, consistent with the expression of *pax3a* mRNA (Figs. 6.2A, B and C). We observed subtle differences in the tissues labelled by CNE1 activity in each stable line in

comparison to the Pax3-GFP^{I150} BAC transgenic line. These include the clear loss of transgene activation in a defined hindbrain rhombomere of L1751 derived embryos (Fig. 6.2A) and the lack of a consistent anterior limit of expression shared by each line. We hypothesised that these differences may be a result of a positional effect of reporter integration into the genome and the influence of local regulatory elements upon transgene activity in each line. We noted the weak labelling of migratory neural crest cells at the somite boundaries in embryos from each stable line stained with GFP antibody at 24 hpf. These cells did not appear to actively transcribe *pax3a* mRNA at this stage (Fig. 4.1D), however they were clearly labelled in Pax3-GFP^{I150} transgenic embryos stained for GFP localisation (Fig. 4.7A). These data suggested that the perdurance of fluorophore proteins might be the main contributor to the labelling of these cells. Transverse sections of trunk tissue demonstrated that CNE1 mediated transcription was restricted to the spinal cord in each stable line and was never observed in developing muscle fibres (Figs. 6.2D, E and F). Cells labelled by GFP staining strongly co-localised with DP312 staining in the dorsal half of the spinal cord and did-not extend ventrally beyond the Pax3/7 boundary. These results strongly suggested that CNE1 was specifically active within the Pax3/7 expression domain of the dorsal spinal cord and further strengthened our analysis of transient transgenic embryos.

We sought to independently confirm our findings of CNE1 activity and extend them to other model systems through the use of published datasets documenting the location of putative enhancers in embryonic tissues. One study that was particularly insightful documented the occupancy of the transcriptional co-activator p300, which has been shown to be predictive of enhancers associated with actively transcribed genes due to its role in the recruitment of the basal transcription machinery (Visel et al., 2009). This study was unique within the literature as it attempted to employ a genome wide approach to enhancer identification and subsequently predict the activity of these regions within defined tissues of the mid-gestation mouse embryo. This was achieved by the creation of 3 datasets that detail the occupancy of p300 in dissected preparations of forebrain, midbrain and limb tissue at E11.5. We were aware that the developmental stage of the tissues utilised in this investigation is comparatively later than our zebrafish assay and therefore past the maxima of Pax3 expression in the spinal cord, however Pax3 is highly expressed in the midbrain and limb bud at this stage (Goulding et al., 1991).

Tracks representing the binding profiles of p300 in each individual tissue were accessed via the NCBI GEO browser under the identifier GSE13845. The WIG file of

each track, documenting the location of binding peaks above an estimated false discovery rate of 0.01, was uploaded to the mm9 (NCBI37) release of the mouse genome via the UCSC genome browser and the position of CNE1 was determined by a BLAT search. These data clearly demonstrated that the only statistically significant peak of p300 binding at E11.5 across the mouse Pax3 locus overlapped with CNE1 in the midbrain-derived dataset (Fig. 6.3). We observed a small peak that fell below the false discovery threshold in the forebrain sample, however p300 binding was not associated with CNE1 in limb tissue. These ChIP-seq results further validated our implementation of comparative genomics by identifying CNE1 as a putative enhancer in an independent experiment. The observation that the sole statistically significant peak of p300 binding was found in midbrain tissue added weight to our hypothesis that this region performed a specific role in regulating the expression of Pax3 during CNS development.

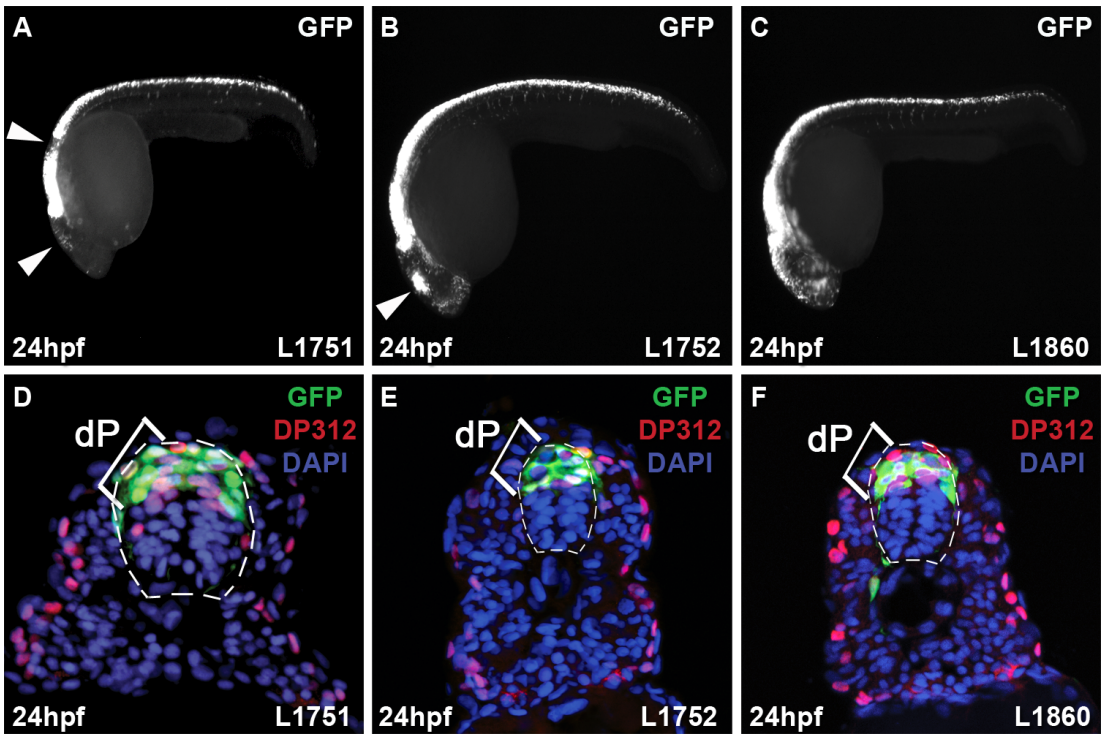


Fig. 6.2 – CNE1 is a CNS specific enhancer

L1751 CNE1 transgenic zebrafish exhibit Citrine transcription within the dorsal spinal cord but lack expression within a single hindbrain rhombomere and the midbrain (arrowheads within A). L1752 transgenics also express Citrine within the dorsal spinal cord and hindbrain, however transcription in the midbrain is restricted to an isolated cluster of cells (arrowhead in B). L1860 transgenic zebrafish appear to recapitulate *pax3a* mRNA expression at 24 hpf (C). Transverse sections of each transgenic line at trunk level demonstrate that CNE1 mediated transcription is restricted to Pax3/7 expressing progenitors in the dorsal neural tube (dP) (D, E and F).

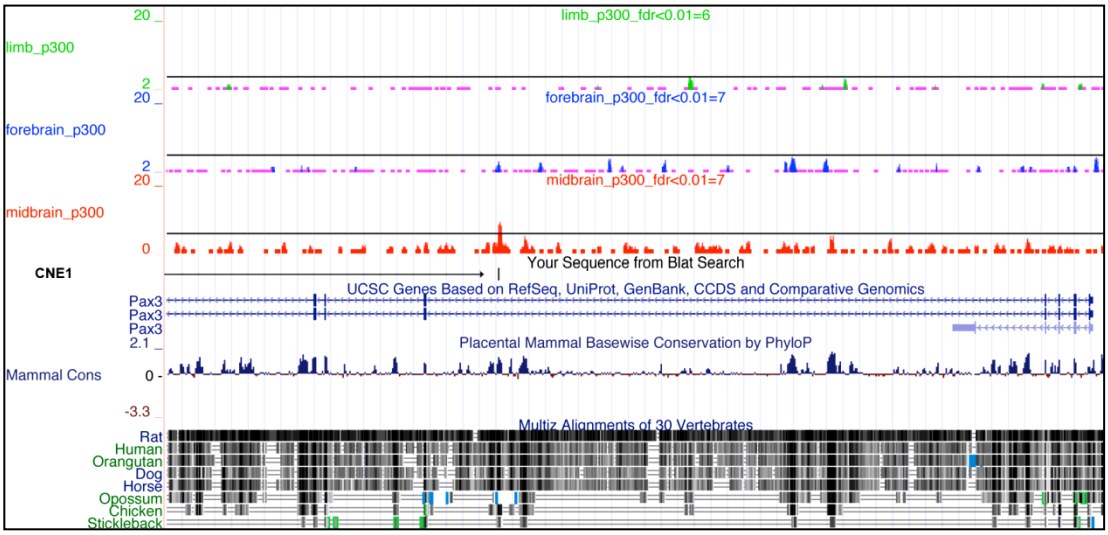


Fig. 6.3 – CNE1 is specifically bound by p300 in CNS derived tissues

UCSC genome browser plot demonstrating the profiles of p300 binding across the Pax3 locus in tissue derived from the mouse limb bud, forebrain and midbrain at E11.5. These data reveal that CNE1 is the only enhancer marked by p300 binding and furthermore, that this regulatory element is specifically active in CNS derived tissues.

ChIP-seq datasets published within Visel et al., 2009.

6.3 – The spatiotemporal profile of CNE1 activity

We sought to further characterise the spatiotemporal dynamics of CNE1 activity *in-vivo* using our stable transgenic zebrafish lines, in order to assess the function of this enhancer in the establishment of the Pax3 expression domain. To this end, we performed a series of wholemount GFP antibody stainings at the developmental stages previously used to determine the expression profile of *pax3a* (Fig. 4.1).

CNE1 driven transgene expression recapitulates *pax3a* mRNA during the early stages of CNS patterning. Accordingly, GFP staining was first observed in dorsal views of the anterior and lateral neural plate at bud stage (Fig. 6.4A). Citrine localisation became restricted to the lateral edges of the intermediate and posterior neural plate at the 6 somites stage, consistent with *pax3a* expression. However, we noted that the segmented pattern of *pax3a* expression in the anterior CNS was not evident in CNE1 stable transgenic embryos assessed at this stage (Fig. 6.4B). Transgene labelling within the spinal cord closely correlated with the expression of *pax3a* mRNA at both the 18 somite stage (Fig 6.4C) and 24 hpf, as previously described (Fig 6.4D). Citrine localisation within CNE1 transgenics differed from *pax3a* expression at later stages of development, possibly due to the perdurance of protein within cells. This effect was particularly evident at 32 hpf, as robust GFP staining was observed in the spinal cord at a stage when *pax3a* transcription is downregulated (Fig. 4.6E). Similarly, GFP staining was retained in tissues of the anterior CNS that do not actively transcribe *pax3a* at 48 hpf (Fig. 4.6F). We noted that the number of cells marked by GFP staining in the spinal cord was greatly reduced within embryos at 48 hpf, supporting our hypothesis that the perdurance of fluorophore protein was the main contributor to these differences. We reasoned that this issue could be avoided by assessing Citrine transcription rather than protein distribution, however we did not perform this experiment as the data acquired at earlier stages was sufficient to establish a role for CNE1 in the induction and maintenance of Pax3 expression. A series assaying CNE1 activity in transient transgenic embryos demonstrated a similar spatiotemporal profile throughout CNS development (data not shown).

The results from our developmental series of CNE1 stable transgenic embryos suggested that the onset of transgene activation coincided with the induction of *pax3a* transcription in the cells of the lateral neural plate. We sought to further analyse this point by preparing cross sections of the intermediate neural plate at bud stage and determining the cellular localisation of Citrine protein relative to Pax3/7. This analysis

confirmed that GFP staining was restricted to the intermediate and lateral regions of the open neural plate (n=9) (Fig. 6.5A), co-localising with the majority of Pax3/7 expressing cells marked by DP312 staining (Fig. 6.5B and C). Sections of anterior regions of the CNS demonstrated that cells labelled by GFP were dorsally located within the tissue and extended towards the medial limit of the neural plate, whereas DP312 staining is not spatially restricted (Data not shown). Taken together, these data strengthened our hypothesis that CNE1 functions to co-ordinate transcription specifically within the Pax3/7 domain of the developing CNS. Furthermore, the observation that Citrine protein was localised within the cells of the lateral neural plate as they initiated Pax3 transcription suggested that this regulatory element might play a role in this process.

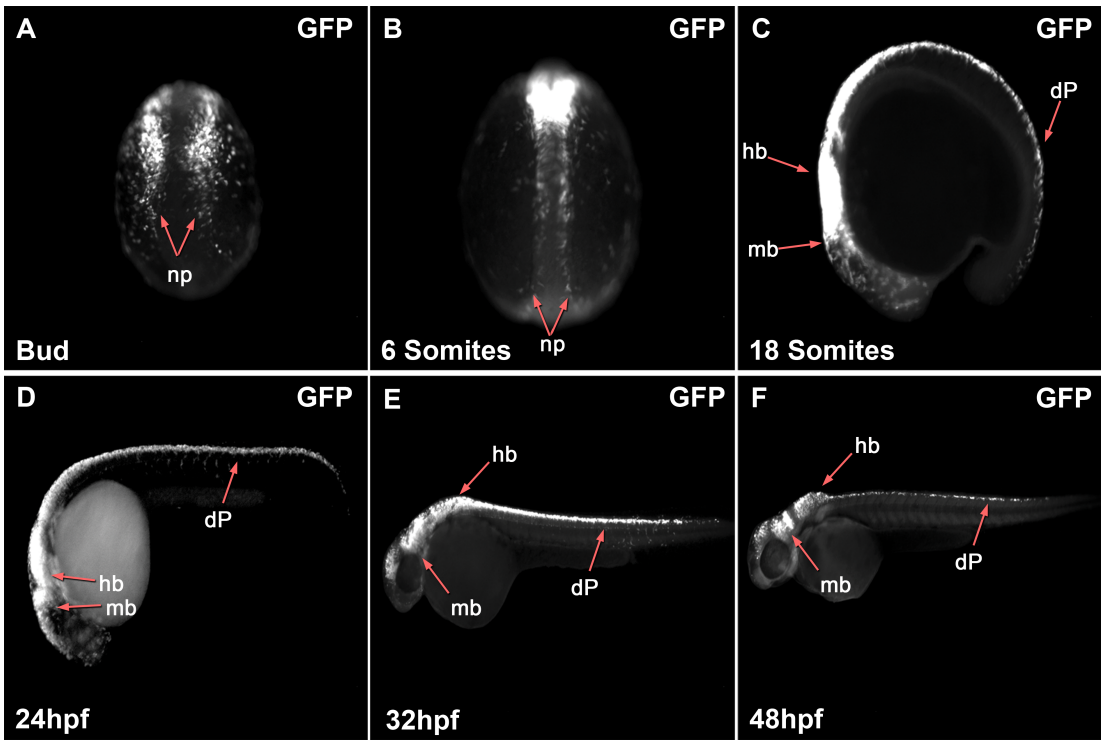


Fig. 6.4 –CNE1 activity recapitulates *pax3a* expression during early development

Dorsal views of CNE1 transgenic zebrafish at bud stage reveal Citrine labelling within the intermediate and lateral regions of the posterior neural plate (A). GFP staining becomes restricted to the lateral neural plate in dorsal view of embryos assessed at the 6 somites stage (B). In lateral view taken at 18 somites and 24 hpf, Citrine protein is expressed within the midbrain (mb), hindbrain (hb) and dorsal progenitors (dP) within the spinal cord (C and D). The perdurance of Citrine protein within cells previously labelled by CNE1 transcriptional activity might account for the observed GFP staining in the spinal cord at 32 hpf and 48 hpf in the absence of *pax3a* transcription (E and F).

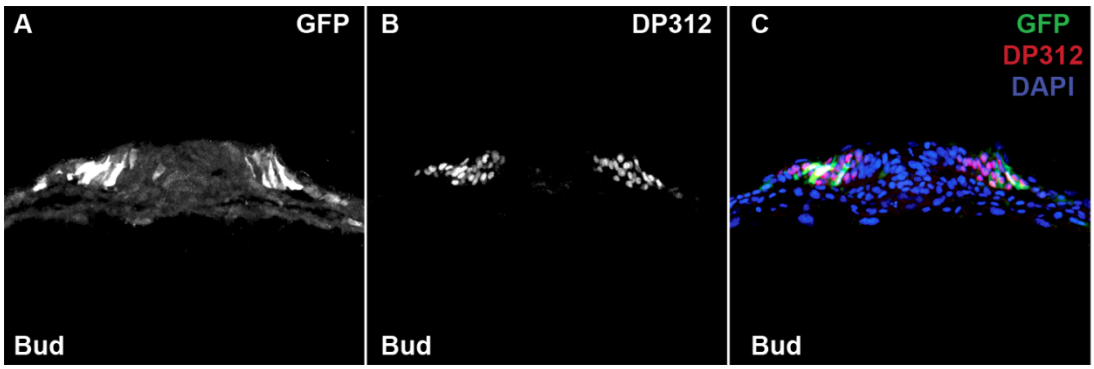


Fig. 6.5 – CNE1 is active in Pax3 expressing cells within the neural plate

Transverse sections of CNE1 transgenic zebrafish at bud stage reveal GFP staining within the intermediate and lateral regions of the neural plate (A). Co-labelling with DP312 demonstrates that CNE1 activity labels Pax3 expressing cells (B and C).

6.4 – Zebrafish CNE1 retains its function in chick

Our analysis of the transcriptional activity mediated by Pax3-CNE1 suggested that it might play a role in both the establishment and maintenance of the Pax3 expression domain during the early stages of CNS development. Based upon these results we chose to investigate this regulatory region further by performing a series of targeted sequence manipulations, with the aim of describing its regulatory potential in terms of the complement of transcription factors it binds. We reasoned that our transient transgenic zebrafish assay would provide the best means of screening a large array of mutant constructs *in-vivo*, however we were aware that the mosaic expression achieved in this assay could make the interpretation of expression data problematic. In order to alleviate these issues and ensure confidence in our data, we sought to develop a second *in-vivo* assay in which the activity of CNE1 could be determined in relation to an internal control of transgenesis. To this end, we performed a series of pilot experiments in which our transgenic reporter cassette, driven by zebrafish CNE1 sequence, was electroporated into the spinal cords of HH stage 11-12 chick embryos. We co-electroporated a plasmid encoding a nuclear localised LacZ reporter as an internal control of the electroporation efficiency, as it can be easily visualised by antibody staining and does not encode a fluorescent reporter. We employed a monoclonal Pax7 antibody as a marker of the Pax3/7 domain in this assay, as it has been previously demonstrated to be more suitable for use in chick tissue than the Pax3 antibody (Ana Ribero, personal communication).

Transverse sections of chick spinal cord tissue, prepared 24 hours post electroporation (hpe), validated this assay as a secondary means of assessing the *in-vivo* activity of a putative enhancer. These data clearly demonstrated that LacZ staining was distributed across the entire DV axis of the experimental side of the neural tube, however CNE1 transcriptional activity was restricted to neuronal progenitors within the Pax3/7 domain (n=12) (Fig. 6.6A). Sections of spinal cord tissue taken at 48 hpe demonstrated that CNE1 remained active in the Pax3/7 domain during CNS development (n=6) (Fig. 6.6B). We noted that the majority of labelled cells at this stage resided within the mantle zone of the spinal cord, suggesting that they were committed to a neuronal fate.

These data demonstrated that *in-ovo* electroporation was a viable method of screening putative enhancers. We observed that transgene activity in this assay was comparably mosaic to that achieved in zebrafish embryos, however the internal control

of trangenesis provided by electroporation experiments enabled the relative activity of an enhancer to be determined. We were reassured to discover that the zebrafish CNE1 sequence retained its function in chick embryos, given the high degree with which this region is conserved across vertebrate species. Furthermore, we observed that the activity of the zebrafish sequence representing CNE1 scaled according to the size of organism in which it operated. This is particularly striking when one considers that the Pax3/7 domain in zebrafish comprises approximately 10 cells, whereas the same domain in chick embryos contains approximately 110. These data suggested that the inputs received by CNE1 scale, both in terms of the size of the tissue and number of cells it contains. This result was expected prior to performing the experiment, as previous studies in the lab have demonstrated that the spinal cord GRN scales according to tissue size (Ana Ribero and Ani Kicheva. personal communication).

The major limitations of chick electroporation experiments, in comparison to zebrafish injections, are the lack of access to large numbers of embryos, the inability to assess enhancer activity outside of the neural tube and the amount of post-experimental processing required to extract data. With these considerations in mind, we chose to assess our sequence manipulations primarily in whole mount preparations of zebrafish embryos at 24 hpf before utilising 24 hpe sections of chick electroporation experiments to verify our observations.

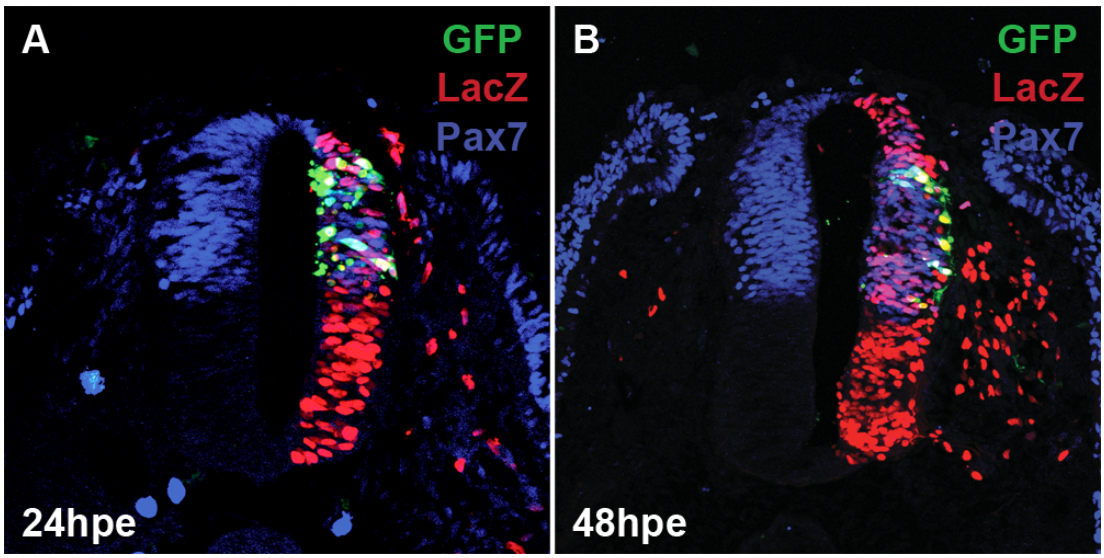


Fig. 6.6 – Zebrafish CNE1 retains activity in the chick neural tube

Co-electroporation of zebrafish CNE1 sequence with nuclear localised LacZ reveals that Citrine expression is restricted to progenitors within the Pax7 expression domain of the chick neural tube at 24 hpe (A). The majority of cells labelled by CNE1 activity reside laterally within the mantle zone in embryos assessed at 48 hpe, suggesting that these cells represent neurons (B).

Chapter 7: Molecular analysis of CNE1 regulation

7.1 – Identifying regulatory motifs within CNE1

In-vivo analysis of Pax3-CNE1 established this region as a functional enhancer, the activity of which recapitulated the expression profile of *pax3a* during spinal cord patterning. To advance this study, we performed a functional dissection of CNE1 to determine the interactions this region mediated between the Pax3 locus and members of the neural tube GRN. To this end, we chose to exploit the selective pressure that exists in functional regulatory elements to identify highly conserved putative TFBS in CNE1. This analysis took advantage of the large evolutionary distances of human-fugu comparisons but extended the span of vertebrates analysed to include mammalian, amphibian, avian, reptilian and fish genomes. Regions of CNE1 homology were identified by cross species BLAST searches and the compiled dataset was aligned using the default parameters of the ClustalW2 algorithm (Larkin et al., 2007).

The aligned data not only confirmed that CNE1 was highly conserved within vertebrate genomes, but also revealed that it contained several clusters of nucleotides that were completely homologous across the evolutionary span of the dataset (Fig. 7.1). We analysed these regions within CNE1 using the MEME suite of sequence analysis tools (Bailey et al., 2009). The default parameters of the algorithm were not suited to the needs of our study, due to the degree of overall homology that existed within the input data, we therefore compensated for this by decreasing the maximum width of discovered motifs from 50 bp to 15 bp, which greatly increased the stringency of the analysis and produced satisfactory results. MEME analysis of CNE1 revealed the presence of 4 highly conserved motifs that correlated with the conserved regions previously identified by ClustalW2 (Table 7.1). We hypothesised that these regions might represent the binding sites that coordinate the transcriptional activity of CNE1 and consequently determine Pax3 expression during CNS development. We sought to investigate the requirement for each motif *in-vivo* by creating individual deletions of these regions within CNE1. This was achieved by mutagenic PCR of the wild type sequence contained in the MiniTol2-Citrine vector using the modified protocol outlined in Table 2.2 and the primers listed in Table 2.3. The activity of each deletion construct was initially assessed by wholemount GFP antibody stainings of transient transgenic zebrafish embryos before transverse sections of electroporated chick embryos were used to verify our observations.

These data revealed the presence of GFP stained cells within the midbrain of all Motif1 deleted transgenic embryos assessed at 24 hpf (n=22) (Fig. 7.2D). CNE1 activity was completely absent from neural progenitor cells within the spinal cord at trunk level, however GFP labelled cells were observed in the most posterior region of 36% of embryos (8/22). Data from chick experiments were consistent with this observation, as only 13% of the embryos marked by LacZ staining demonstrated transgene activity within spinal cord sections at 24 hpe (n=1/8) (Fig. 7.2E). 95% of zebrafish embryos injected with the Motif2 deletion construct exhibited GFP staining throughout the dorsal CNS, in a profile consistent with the wild type enhancer (n=37/38) (Fig. 7.2F). Furthermore, 66% of LacZ positive chick spinal cords demonstrated transgene activation within the dorsal Pax3/7 expression domain (n=4/6) (Fig. 7.2G). A decrease in transgene induction throughout the entire AP axis of the CNS was observed in Motif3 deleted fish, in comparison to the activity of the wild type enhancer (n=26) (Fig. 7.2H). Mosaic clusters of GFP positive cells were present within the dorsal trunk of 77% (20/26) of these embryos, however these cells were located at the dorsal limit of the spinal cord and did not exhibit a characteristic neural progenitor morphology. We did not observe any transgene expression in the LacZ positive spinal cord sections of chick embryos electroporated with the Motif3 deletion construct, confirming our analysis of zebrafish experiments (n=0/5) (Fig. 7.2I). Deletion of Motif4 appeared to result in a decrease in the activity of CNE1, as only 40% of transgenic fish exhibited GFP staining within the spinal cord (n=8/20). However, we noted that the distribution of Citrine expressing cells within the CNS of these samples was comparable to the wild type enhancer (Fig. 7.2J). This observation was verified by electroporation of the Motif4 deletion construct, which resulted in transgene activation within the dorsal spinal cord of 80% of chick embryos (n=4/5) (Fig. 7.2K). These data are plotted and statistically analysed within Fig. 7.3

Taken together, the data from our series of Motif deletion experiments suggested that Motif1 was specifically required for the correct activity of CNE1 within spinal cord progenitor cells whereas loss of Motif3 reduced the transcriptional activity of the enhancer throughout the anterior-posterior axis of the CNS. The function of CNE1 was largely retained in the context of Motif2 and 4 deletions, suggesting that these regions were not critical for the activity of this regulatory element.

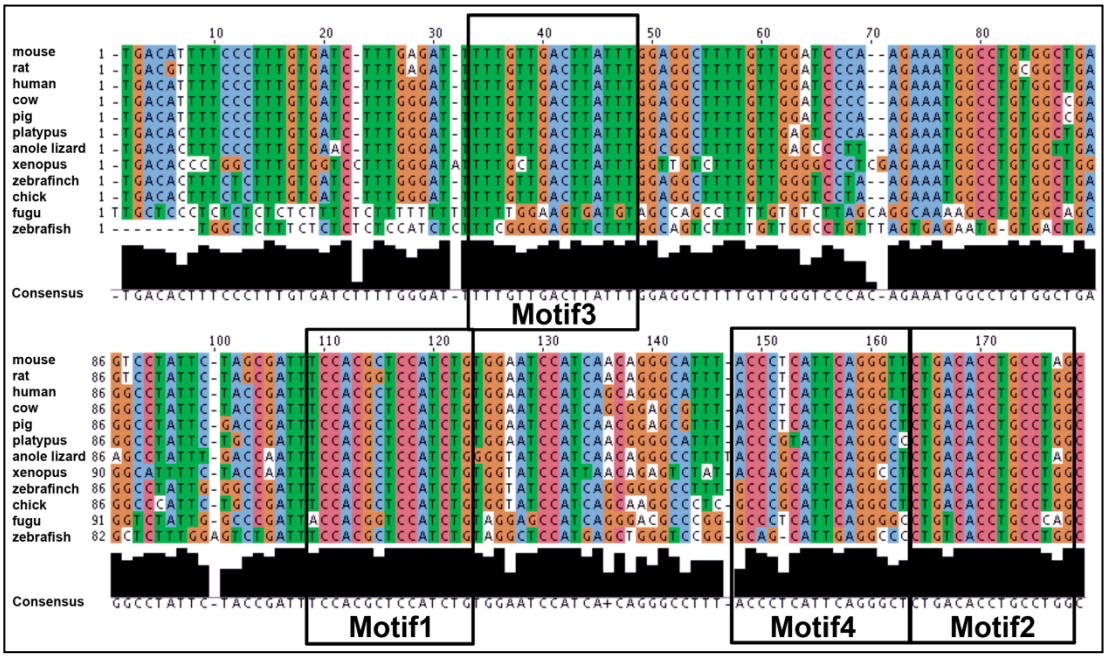


Fig. 7.1 – Clusters of nucleotides within CNE1 are conserved across the span of vertebrate evolution

ClustalW2 alignment of CNE1 sequences extracted from 12 vertebrate genomes reveals blocks of nucleotides that are perfectly conserved. These regions represent putative TFBS that might underlie the enhancer activity of CNE1 *in-vivo*.

The motifs defined in Table 7.1 are highlighted in the alignment for clarity.

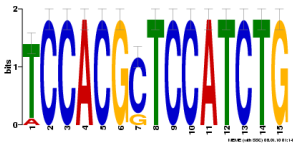
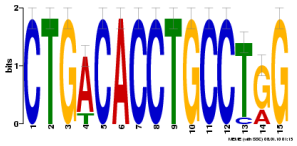
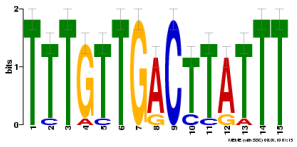
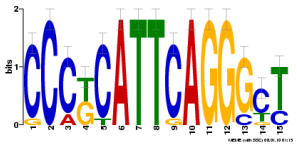
Name	Logo (12 vertebrate genomes)	Location within CNE1 (zebrafish)	Location within Fig. 7.1
Motif1		Bases 99 – 114	Position 109-123
Motif2		Bases 152 -167	Position 164-178
Motif3		Bases 48-63	Position 34-48
Motif4		Bases 139-154	Position 148-163

Table 7.1 – Identification of putative regulatory motifs within CNE1

MEME analysis of CNE1 sequences extracted from 12 vertebrate genomes. Statistically over-represented 15 bp motifs, their location within the zebrafish sequence of CNE1 and within Fig. 7.1 are listed according to their conservation.

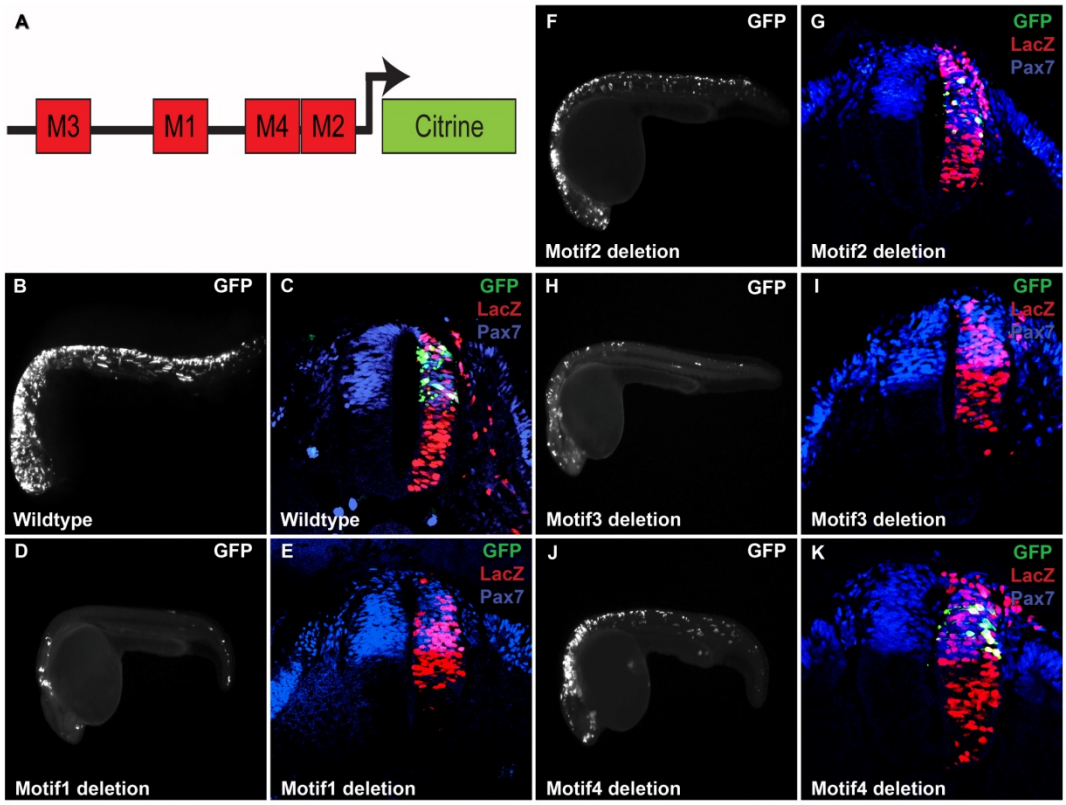


Fig. 7.2 – Motif1 and Motif3 are required for CNE1 function *in-vivo*

Schematic detailing the organisation of motifs within CNE1 (A). The wildtype CNE1 enhancer labels cells in the dorsal neural tube of both zebrafish (B) and chick (C) embryos, assessed at 24 hpf and 24 hpe, respectively. LacZ staining in chick embryos demonstrates the efficiency of electroporation and Pax7 staining marks the Pax3/7 domain. Motif1 deletion results in a loss of CNE1 activity in the spinal cord progenitors of the zebrafish neural tube (D). A similar loss of CNE1 activity is observed in chick embryos electroporated with the Motif1 deletion construct (E). Motif2 is not required for CNE1 activity in either zebrafish (F) or chick embryos (G). Deletion of Motif3 results in complete loss of CNE1 activity in spinal cord progenitors in both zebrafish (H) and chick embryos (I). In contrast, deletion of Motif4 does not affect the spatial distribution of Citrine labelling in the dorsal neural tube (J and K).

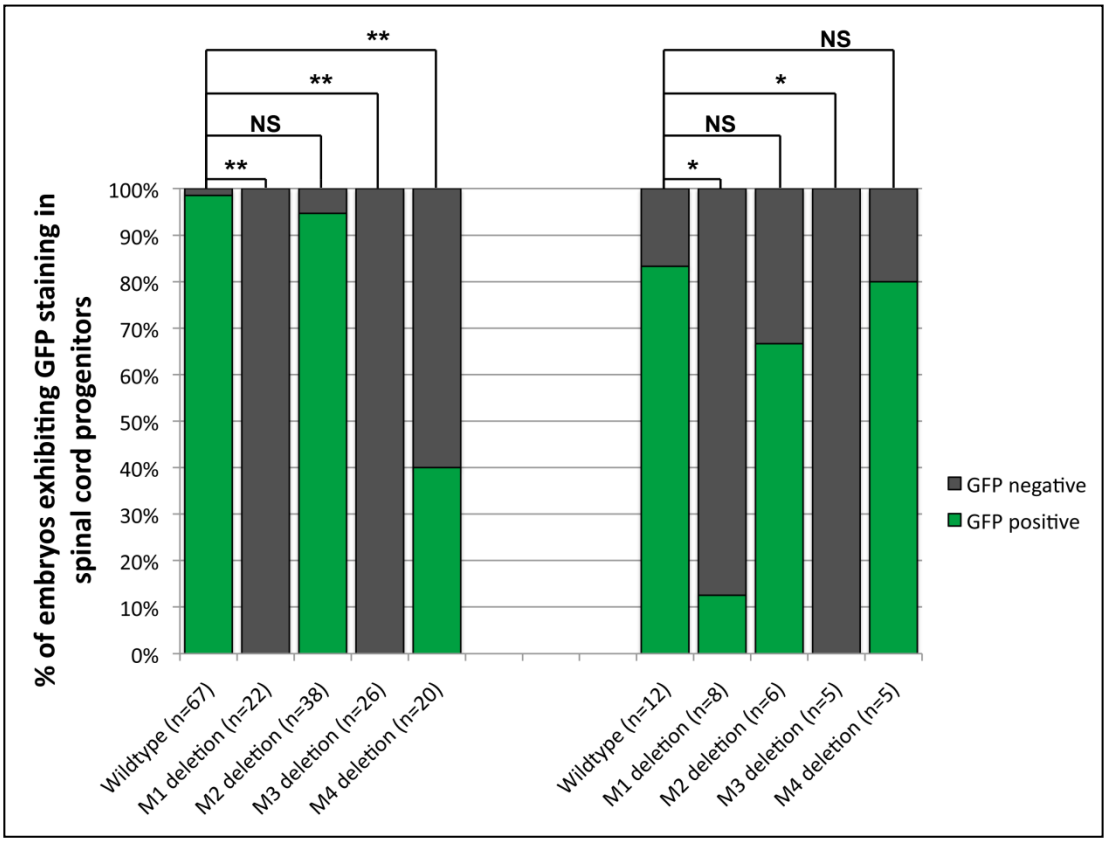


Fig. 7.3 – Quantification of the effects of motif deletion upon CNE1 activity

The percentages of transgenic zebrafish embryos exhibiting GFP staining in spinal cord progenitor cells (within the span of the yolk extension) assessed at 24 hpf is plotted. Similarly, the percentages of electroporated chick embryos exhibiting GFP staining in transverse sections of brachial to lumbar sections of spinal cord is provided. These data demonstrate that loss of Motif1, 3 and 4 results in a highly significant loss of CNE1 activity in zebrafish spinal cord progenitors. In constrast, data obtained from chick experiments indicates that loss of Motif4 does not effect CNE1 activity. These data confirm that Motif1 and Motif3 are required for the enhancer activity of CNE1 *in-vivo*.

Two-tail Fisher's exact test (* = $P<0.005$, ** = $P<0.001$ and NS = not significant).

7.2 – Functional dissection of CNE1

We aimed to utilise the data created by MEME analysis of CNE1 to quantitatively assess the similarity of Motifs 1 and 3 to defined transcription factor binding matrices. This was achieved by comparing the desired PSPM, representing the conservation of each position within the motif, to defined matrices in the JASPAR (Sandelin et al., 2004a) and UniPROBE (Newburger and Bulyk, 2009) databases using the default parameters of the TomTom motif analysis tool (Gupta et al., 2007) (Table 7.2). Further analysis of the cross species consensus sequence surrounding each motif was performed using the ConSite (Sandelin et al., 2004b) and JASPAR databases. This revealed the presence of binding sites for Hox, Tead and Pbx factors within the 10 bp immediately downstream of Motif1 (Table 7.2). These candidates were added to our list of putative CNE1 regulators as they have previously been demonstrated to bind upstream elements associated with the Pax3 locus (Milewski et al., 2004; Pruitt et al., 2004).

We were particularly intrigued by the identification of a 14 bp putative Pax6 binding site within Motif1 by TomTom (Fig. 7.4), as members of the Pax gene family are known to participate in selective auto and inter-regulatory interactions (Frost et al., 2008). Analysis of Motif1 revealed that it did not contain the palindromic site known to be preferentially occupied by paired type homeodomains (Soleimani et al., 2012 ; Wilson et al., 1993), which raised the possibility that this sequence represented a variant of the PD binding site. We tested this hypothesis by determining the homology shared between Motif1 and published definitions of PD binding sites occupied by Pax6 (Epstein et al., 1994), Pax5 (Adams et al., 1992) and *paired* itself (Treisman et al., 1991) using ClustalW2. These data indicated that the 5' region Motif1 shared a high degree of homology to the selected PD binding sites and furthermore that all of the input sequences were identical over bases 4-8 (Fig. 7.3). The 3' region of the alignment diverged and poor conservation was observed between Motif1 and the Pax6 PD binding site. This result was consistent with the hypothesis that Motif1 represented a PD binding site, however it also revealed that the site occupied by Pax6 was not identical to this sequence. Our *in-vivo* assessment of the requirement for Motif1 within CNE1 suggested that this region facilitated the binding of a positive regulatory input that was required for enhancer function in the dorsal spinal cord. Our analysis of the putative PD binding site in Motif1, in addition to the fact that Pax6 is not expressed in the majority of cells in which CNE1 is active, suggested that this region might not represent a functional

Pax6 binding site. Instead, we speculated that Motif1 might represent a PD binding site that could mediate the binding of Pax3/7 proteins to the Pax3 locus.

A site directed mutagenesis screen was performed within Motif1, based upon the incidence of annotated TFBS within this sequence. We previously observed that the majority of transcription factors were associated with two distinct binding sites, a 5' RBPJ site and a 3' E-box, both of which were targeted for mutagenesis using the primers listed in Table 2.4. The Tead binding site associated with Motif1 was also selected for manipulation, as it partially overlapped the sequence deleted in previous experiments. We chose not to specifically target the putative PD binding site within Motif1, as we reasoned that it would be disrupted by any mutation made within this sequence.

Mutation of the RBPJ site did not recapitulate the effect of Motif1 deletion, as 88% of fish embryos demonstrated transgene labelling throughout the CNS at 24 hpf (n=22/25) (Fig. 7.5D) and 75% of chick embryos exhibited GFP staining within the dorsal spinal cord (n=6/8) (Fig. 7.5E). Similar results were obtained in Tead binding site manipulations, as 88% (n=29/33) of TeadMut1 and 98% (n=39/40) of TeadMut2 injected zebrafish embryos retained Citrine labelling throughout the CNS (Figs. 7.5F and H). These results were verified in sections of chick spinal cord tissue, in which we observed GFP staining in 80% of TeadMut1 (n=4/5) and 88% of TeadMut2 (n=7/8) electroporated embryos (Figs. 7.5G and I). These data are plotted in Fig. 7.6.

The E-box sequence within Motif1 was targeted by combining two published manipulations, each of which has been shown to be sufficient to preclude the binding of bHLH proteins to the site (Henke et al., 2009 ; Thattaliyath et al., 2002). All GFP positive fish embryos injected with the E-boxMut1 construct demonstrated transgene labelling within the tissues of the anterior CNS at 24 hpf (n=48/48) (Fig. 7.7D). We frequently observed GFP staining in large dorsally located cells within the trunk, however we did not consider these cells to be neural progenitors as they appeared to reside outside of the spinal cord and did not exhibit the characteristic progenitor morphology. 13% of chick embryos electroporated with the E-boxMut1 construct retained GFP labelling within the spinal cord (n=1/8) (Fig. 7.7E). These data suggested that the E-box binding site within the 3' region of Motif1 was required for the correct activity of CNE1 and furthermore, that the effect of mutating this site recapitulated the deletion of Motif1 *in-vivo*. We chose to further investigate the requirement of this site within Motif1 by creating individual mutations in the 5' (E-boxMut2) and 3' (E-boxMut3) regions and assaying their effects on the transcriptional output of CNE1.

Transgene induction within spinal cord progenitor cells was observed in 8% of E-boxMut2 injected zebrafish (n=2/26) (Fig. 7.7F) and 33% of electroporated chick embryos (n=2/6) (Fig. 7.7G). In contrast, Citrine labelling was not seen in spinal cord progenitors of E-boxMut3 injected fish (n=0/19) (Fig. 7.7H) and was retained in only 20% of chick embryos (n=1/5) (Fig. 7.7I). These data are plotted in Fig. 7.8.

Taken together, these data suggested that the requirement of Motif1 for the activity of CNE1 was not dependent upon the integrity of either the 5' RBPJ or downstream Tead binding sites. A large mutation of the 3' E-box binding site reduced CNE1 activity in a manner that was consistent with the effect of deleting the entire 15 bp motif. Further dissection of this site suggested that E-boxMut3, representing a 2 bp substitution in the sequence of CNE1, also recapitulated the deletion of Motif1. These data suggested that transcription factors bound CNE1 via the E-box in Motif1 to activate or maintain transcription within the spinal cord and furthermore, that the 3' region of the site was dominant in the context of this interaction.

Sequence manipulations of the binding sites contained within Motif3 were designed to target the 5' region and synthesised using the primers listed in Table 2.4. This site was predicted to facilitate the binding of Fox, Sox and Lef family members, all of which are known to play key roles in CNS development (Table 7.2). We hypothesised that the binding of these factors could mediate a general positive regulatory input upon CNE1, given the reduced level of transcriptional activation observed throughout the CNS in Motif3 deleted transgenic zebrafish. Injection of Motif3-Mut1 did not reduce GFP staining in anterior CNS tissues, compared to wild type CNE1 (n=41) (Fig. 7.9D). We observed GFP staining within spinal cord progenitors of 61% (n=25/41) of these embryos, however the distribution of the transgene was highly mosaic and limited to small clusters of cells. Electroporation of Motif3-Mut1 resulted in transcriptional activation within the spinal cord of 43% of chick embryos (n=3/7) (Fig. 7.9E). Motif3-Mut2 appeared to have a greater effect upon CNE1 activity in fish embryos and resulted in a general decrease in the level of Citrine produced throughout the CNS that was comparable to the effects of Motif3 deletion (n=32) (Fig. 7.9F). Transcription within spinal cord progenitors was greatly decreased compared to wild type CNE1, as only 19% (n=6/32) of embryos exhibited GFP staining in these cells. 40% of chick embryos electroporated with the Motif3-Mut2 construct retained transgene expression within the Pax3/7 domain (n=2/5) (Fig. 7.9G). These data are summarised and analysed in Fig. 7.10.

Taken together, our analysis of Motif3 mutant constructs provided further evidence of a general positive regulatory input that was required for CNE1 activity throughout the developing CNS. We hypothesised that this effect may be mediated by the binding of candidate Fox, Sox and Lef family transcription factors to the 5' region of Motif3 and that this interaction was destabilised by our second mutation within this site.

Motif	TomTom Matches (E<10)	Matching Matrices identified by TomTom (JASPAR and UniPROBE)	Surrounding consensus sites (ConSite and JASPAR)
1	17	Hkb, Pax6, Arnt:Ahr, ZNF354C, Ascl2, HSF1, YY1, Irf6, SP1, Sp4, Gata1, Bcl6b, TAL1::TCF3, Tcfe2a, Zfp410, Snail, Bhlhb2	HoxA5, HoxA7, Tead1, Pbx1
3	25	Foxa2, Foxj3, tll, Mafb, Foxk1, Foxl1, FOXD1, FoxA1, HCM1, fkh, FKH2, Foxj1, FKH1, br_Z4, Sox1, Mafk, Tcf1, RORA, Evi1, slp1, AP1, Yap7	N/A

Table 7.2 – Phylogenetically conserved TFBS associated with CNE1-Motif1 and Motif3

TomTom analysis of the PSPMs representing Motif1 and Motif3 reveals that these regions consist of multiple, highly conserved TFBS. These data are supplemented with *in-silico* annotation of the sequence surrounding Motif1, revealing binding sites for several transcription factors previously implicated in the regulation of Pax3 expression.

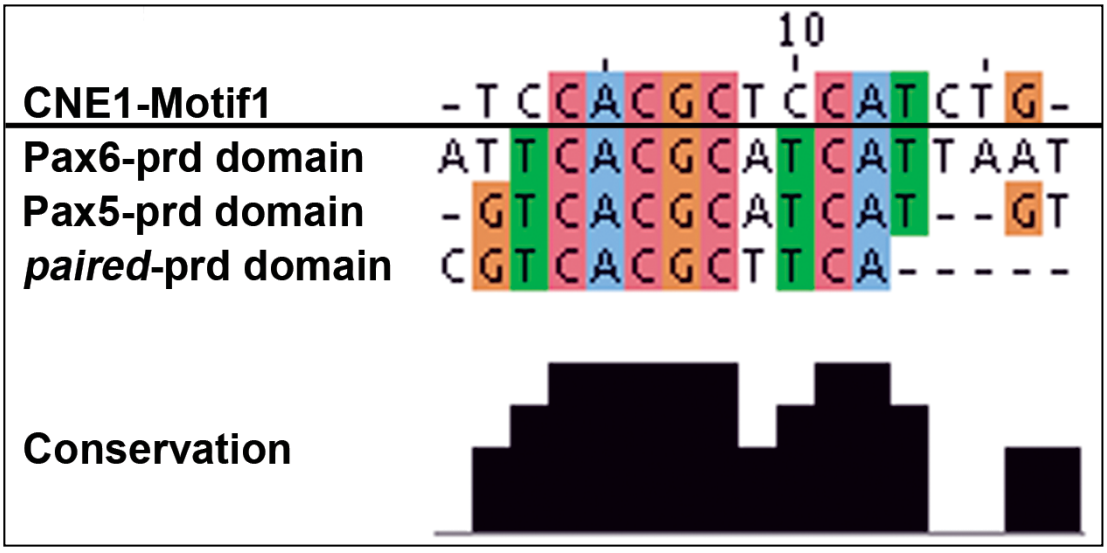


Fig. 7.4 – CNE1-Motif1 represents a putative PD binding site

Alignment of CNE1-Motif1 with published definitions of the Pax6, Pax5 and *paired* PD binding sites using ClustalW2. These data demonstrate consenus between all input sequences over bases 4-8, whereas poor conservation is observed in the 3' region of the alignment.

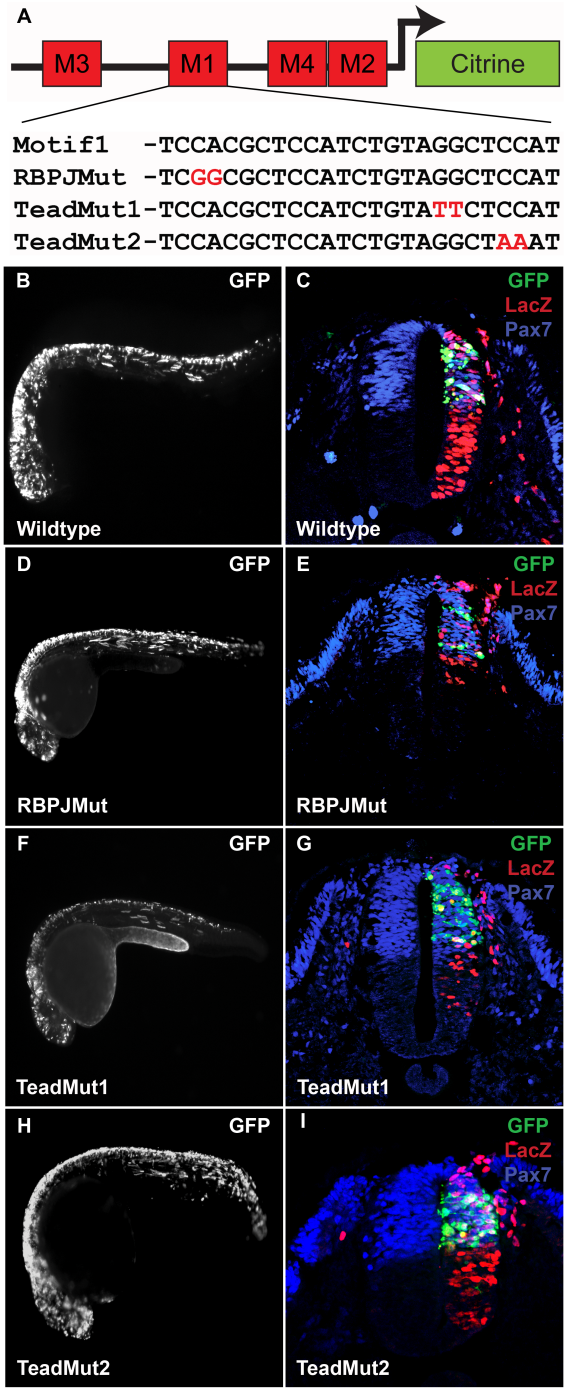


Fig. 7.5 – Mutation of the RBPJ or Tead sites associated with Motif1 does not affect the transcriptional activity of CNE1

Schematic highlighting the mutations induced within the RBPJ and Tead binding sites associated with Motif1 (A). Mutation of the RBPJ site does noes result in a loss of CNE1 activity in the spinal cord of zebrafish (D) and chick embryos (E), compared to wildtype (B and C). Similarly, two individual mutations designed to target the Tead binding site located immediately 3' to Motif1 do not affect the distribution of GFP staining *in-vivo* (F, G, H and I).

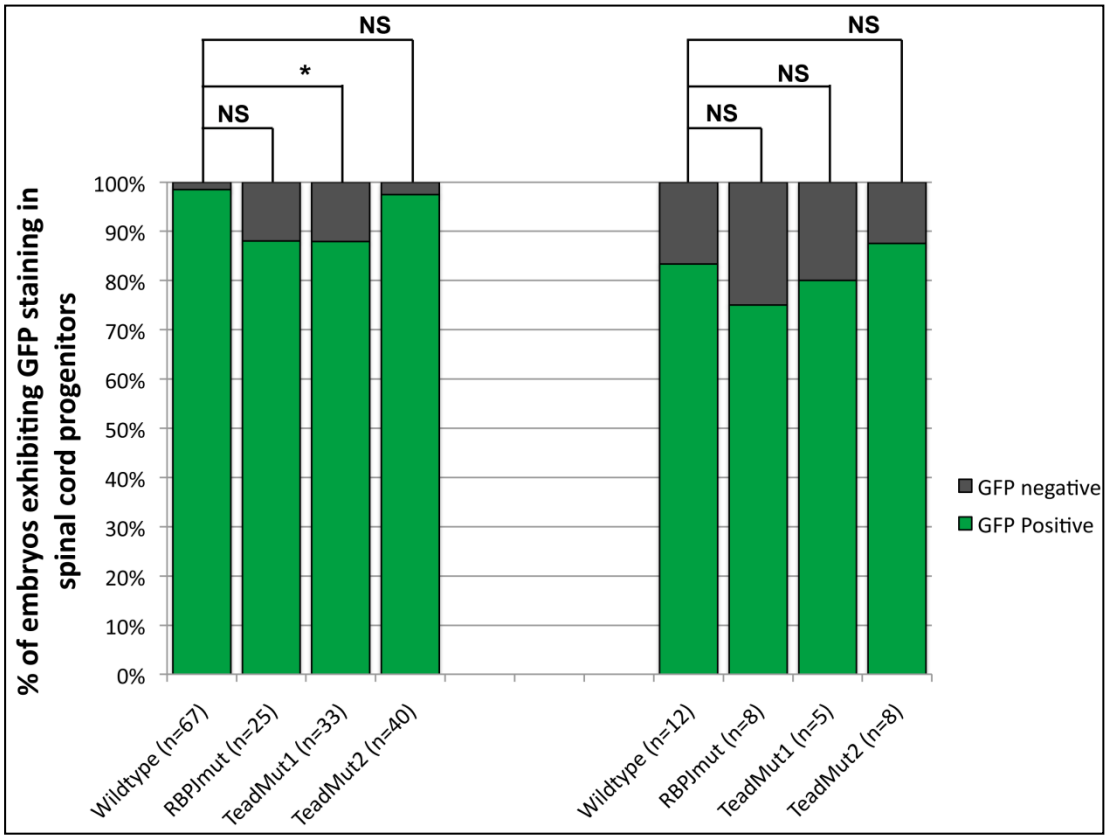


Fig. 7.6 - Quantification of the effects of RBPJ and Tead site mutations upon CNE1 activity

The percentages of transgenic zebrafish embryos exhibiting GFP staining in spinal cord progenitor cells (within the span of the yolk extension) assessed at 24 hpf is plotted. The percentages of electroporated chick embryos exhibiting GFP staining in transverse sections of brachial to lumbar sections of spinal cord is also provided. These data demonstrate the mutation of the RBPJ site within Motif1 does not significantly alter CNE1 activity. 88% of TeadMut1 transgenic zebrafish exhibit GFP staining in spinal cord progenitors, however this result is significantly different from the wildtype control. The mutation of this binding site does not produce a significant reduction in CNE1 activity when assessed in chick embryos. TeadMut2 demonstrates no effect upon CNE1 activity *in-vivo*.

Two-tail Fisher's exact test (* = $P<0.005$ and NS = not significant).

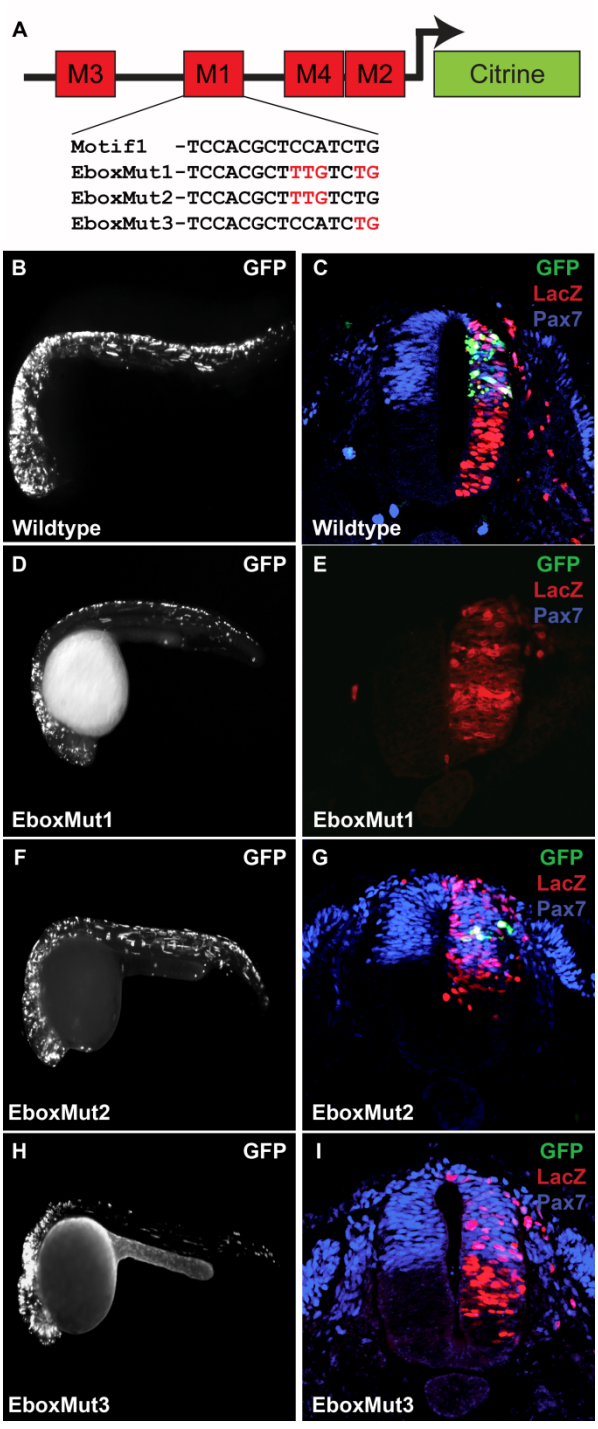


Fig. 7.7 – The 3' region of Motif1 is required for CNE1 activity

Schematic outlining the mutations induced within the E-box of Motif1 (A). E-boxMut1results in a loss of CNE1 activity in spinal cord progenitors of both zebrafish and chick embryos (D and E) compared to th wildtype enhancer (B and C). E-boxMut2 strongly reduces CNE1 activity in progenitor cells within the zebrafish neural tube (F), however Citrine labelling is retained in 60% of chick embryos assessed (G). Consistent with the effects of Motif1 deletion, E-boxMut3 attenuates GFP staining in zebrafish neural tube progenitors (H) and reduces enhancer activity in chick embryos (H and I).

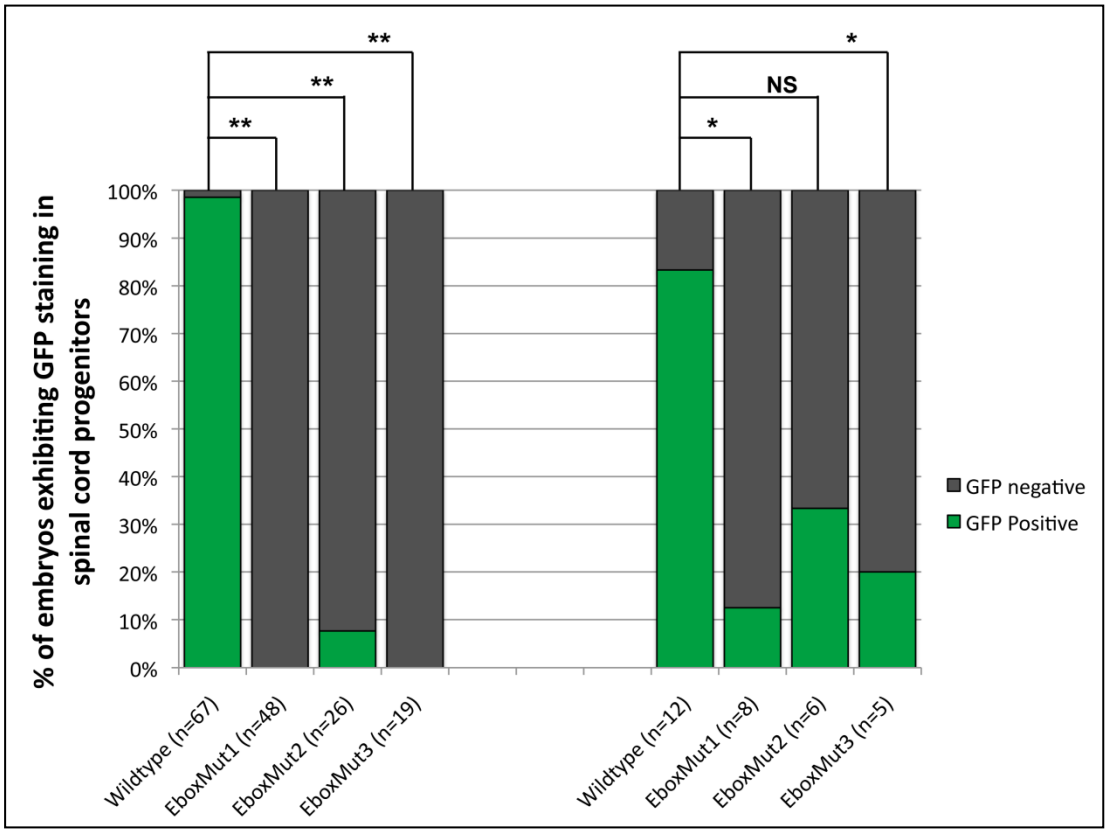


Fig. 7.8 - Quantification of the effects of E-box mutations upon CNE1 activity

The percentages of transgenic zebrafish embryos exhibiting GFP staining in spinal cord progenitor cells (within the span of the yolk extension) assessed at 24 hpf is plotted. In addition, the percentages of electroporated chick embryos exhibiting GFP staining in transverse sections of brachial to lumbar sections of spinal cord is also provided. These data demonstrate that all of the mutations induced with the E-box of Motif1 result in a highly significant reduction of CNE1 activity in zebrafish embryos. In chick, EboxMut1 and EboxMut3 produces a loss of CNE1 activity whereas EboxMut2 is not statistically different from the control experiment. These data are in agreement with the conclusion that the 3' region of the E-box binding site in Motif1 is required for CNE1 activity.

Two-tail Fisher's exact test (** = P<0.001, * = P<0.005 and NS = not significant).

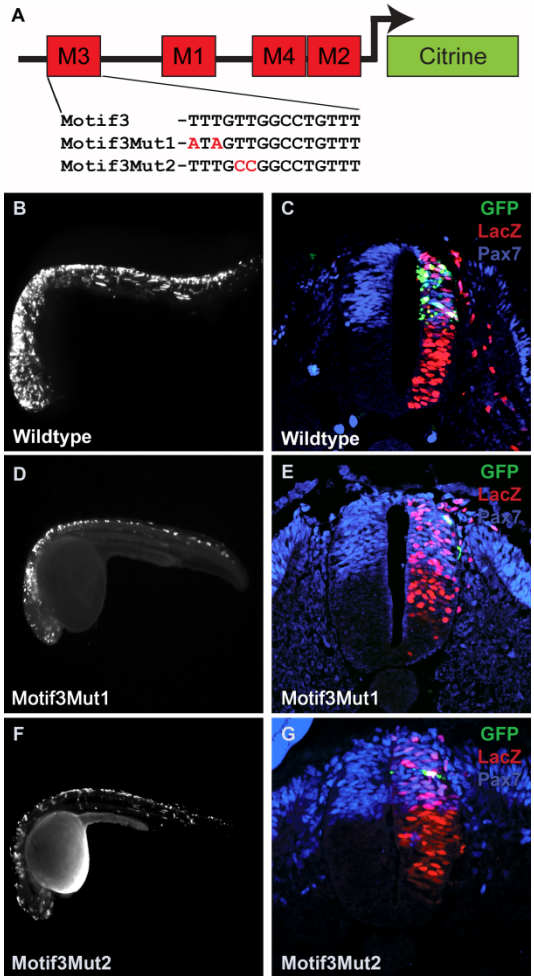


Fig. 7.9 – Mutation of the HMG box within Motif3 reduces CNE1 activity *in-vivo*

Schematic outlining the mutations induced within the 5' region of Motif3 (A). 61% of Motif3^{Mut1} transgenic zebrafish exhibit GFP staining within spinal cord progenitor cells (D). 43% of chick embryos retain Citrine labelling within the Pax3/7 domain of the neural tube (E). Only 19% of Motif3^{Mut2} transgenic zebrafish exhibit CNE1 activity in neural tube progenitors (F), however 40% of chick embryos retain GFP staining in the dorsal neural tube (G)

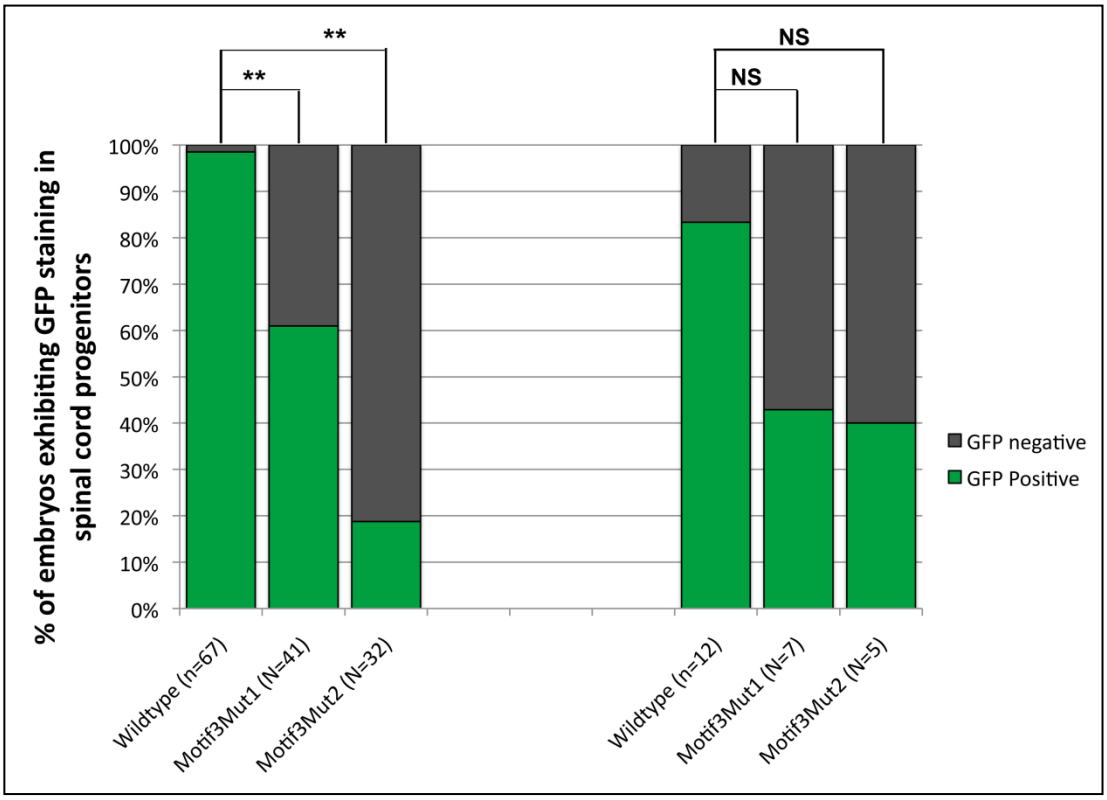


Fig. 7.10 - Quantification of the effects of Motif3 mutations upon CNE1 activity

The percentages of transgenic zebrafish embryos exhibiting GFP staining in spinal cord progenitor cells (within the span of the yolk extension) assessed at 24 hpf is plotted with the percentages of electroporated chick embryos exhibiting GFP staining in transverse sections of brachial to lumbar sections of spinal cord. These data demonstrate that both mutations induced within the 5' region of Motif3 result in a highly significant reduction in CNE1 activity in zebrafish embryos. These sequence manipulations also produce a consistent loss of CNE1 activity in chick embryos, however these data are not statistically different from control experiments.

Two-tail Fisher's exact test (** = $P<0.001$ and NS = not significant).

7.3 – Motif1 is a functional paired domain binding site

In-vivo functional dissection of Pax3-CNE1 resulted in the identification of two conserved motifs that were required for the activity of this enhancer. Furthermore, we were able to describe the requirement of these motifs in terms of the integrity of the discrete binding sites they contain. We aimed to build upon our *in-vivo* analysis of CNE1 regulation by assessing the binding of candidate transcription factors to CNE1 *in-vitro* and investigating the effects of our sequence manipulations upon these interactions. This was achieved in a series of EMSAs, a technique commonly used to investigate the binding of transcription factors to a defined fragment of DNA. Our initial experiments were designed to assess the ability of *in-vitro* synthesised proteins to bind 33 bp radioactively labelled double stranded DNA probes, centred upon the desired regions of CNE1 within the zebrafish genome. The oligonucleotides and protocols used to create each EMSA probe are described in Chapter 2.3 and Table 2.6.

Our *in-silico* analysis of CNE1-Motif1 indicated that it contained, or was proximal to, putative binding sites for HoxA5, HoxA7, Pbx1 and Tead1. We chose to select these candidates for EMSA analysis as previous studies have implicated members of these transcription factor families in the regulation of Pax3 expression (Milewski et al., 2004; Pruitt et al., 2004). We synthesised Tead2 protein rather than Tead1 in order to maintain consistency with this data (Milewski et al., 2004). Both Pax3 and Pax7 proteins were selected for analysis in these experiments to test our hypothesis that Motif1 represented a binding site for the PD encoded by these transcription factors. Six2 protein was synthesised and used as a negative control in this experiment, as it was not predicted to bind the sequence contained within the Motif1 EMSA probe. The protocols used to synthesise proteins and perform this assay are outlined in Chapter 2.3. The results of this experiment suggested that neither Tead2, Pbx1, HoxA5 nor HoxA7 proteins were able to bind the Motif1 DNA probe, as the distribution of complexes observed in the autoradiograph was identical to those present in the negative control reaction (Fig. 7.11). In contrast, both Pax3 and Pax7 bound the radiolabelled probe to form DNA-protein complexes that were both larger and more abundant than those present in the control experiment. Furthermore, the Motif1-Pax7 interaction appeared to produce two distinct complexes, which is commonly observed if transcription factors are able to dimerise upon DNA (Fig. 7.11).

These data verified our annotation of Motif1 as a putative PD site that was sufficient to facilitate the binding of both Pax3 and Pax7. We investigated this point

further by performing an implementation of the EMSA that is based upon the interaction of total protein, also known as nuclear extract, with DNA. The protocol used to perform nuclear extractions of chick tissue is outlined in Chapter 2.3. DNA-protein interactions were identified in this assay by the ability of specific antibodies to interact with candidate transcription factors within the nuclear extract, resulting in the formation of a larger or 'supershifted' complex following gel electrophoresis. The positive control reaction for this experiment, in which Motif1 DNA was bound by nuclear extract in the absence of antibodies, resulted in the formation of a complex as expected (Fig. 7.12). Addition of a mixture containing monoclonal antibodies raised against Pax3 and Pax7 resulted in a shift of the distribution of labelled probe to a larger complex, indicating that these proteins were present within the nuclear extract and bound to DNA (Arrow in Fig. 7.12). A supershift was also produced by the addition of DP312 antibody to the reaction, which we have previously demonstrated to specifically recognise the paired type homeodomains of Pax3 and Pax7 *in-vivo* (Arrow in Fig. 7.12). Addition of a monoclonal Pax6 antibody did not result in a shift of the DNA-protein complex, suggesting that the PD binding site in Motif1 was specifically occupied by both Pax3 and Pax7. This result was consistent with our analysis of DP312 antigen recognition *in-vivo* and our *in-silico* analysis of the PD site represented by Motif1. The final lane in the autoradiograph documents the results of a negative control reaction containing labelled probe in the absence of protein, which was used to determine the presence of non-specific experimental artefacts.

Functional dissection of Motif1 in-vivo revealed that the integrity of the E-box binding site was required for the correct activity of CNE1. We hypothesised that mutation of this region may weaken the interaction between Pax3 and Pax7 proteins and DNA, providing a molecular basis to our observations. We tested this prediction by performing a series of EMSAs in which the wild type Motif1 probe was placed in competition for Pax7 binding with non-labelled DNA probes. The control series of reactions in this experiment clearly demonstrated that non-labelled Motif1 probe effectively bound Pax7 protein (Fig. 7.13). This was visualised as the progressive loss of signal from the complex once the amount of non-labelled DNA exceeded 5ng/reaction, which marked the point of equilibrium between the two probes (asterisks in Fig. 7.13). We employed this experimental paradigm to assess the ability of E-boxMut3 DNA to compete for Pax7 binding against the wild type Motif1 probe. The double stranded DNA probe used in these experiments was created by annealing the oligonucleotides previously employed in mutagenic PCR. Titration of Ebox-Mut3 DNA

to a four fold excess in the reaction did not significantly reduce the binding of Pax7 to wild type Motif1, as demonstrated by the persistence of complex formation in the presence of 20ng of competitor probe (Fig. 7.13). These data indicated that the 2 bp substitution induced by the E-boxMut3 manipulation affected the integrity of the PD site in a manner that reduced its affinity for Pax7 binding.

The PD binding site within CNE1 appeared to be required for the activation of transcription within spinal cord progenitors *in-vivo*, suggesting that the binding of Pax3 and Pax7 to Motif1 represented a positive regulatory input upon the enhancer. We aimed to test this hypothesis by expressing Pax3 and Pax7 throughout chick spinal cord and assessing the ability of these proteins to ectopically induce CNE1 mediated transcription. We chose to use the RCAS retroviral vector to induce the expression of our desired proteins, as we had previously observed that high levels of ectopic Pax3/7 greatly affects neural tube patterning and morphology (data not shown). An oncogenic fusion of Pax3 and FOXO1A, which has been shown to function as an obligate transcriptional activator, was cloned into the Cla1 site of the RCAS vector (Relaix et al., 2003). Electroporation of this construct was sufficient to induce expression of Pax3 within the ventral spinal cord at 24hpe (n=9) (Fig. 7.14A). CNE1 was cell autonomously activated in Pax3 expressing cells within the ventral spinal cord of Pax3-FOXO RCAS co-electroporated embryos (Figs. 7.14B and C). A similar experiment, in which the RCAS vector was used to express the full length Pax7 cDNA, also resulted in the cell autonomous ectopic induction of Citrine transcription within the spinal cord (n=7) (Figs. 7.14D,E and F). These data confirmed our hypothesis that Pax3 and Pax7 bound the PD within CNE1 to induce transcription.

A series of experiments were performed to assess the interaction of candidate transcription factors with Motif3, with the aim of investigating the molecular basis of the reduction in transcription observed throughout the CNS when this region was deleted from CNE1. A 33 bp probe representing Motif3 was used in assays with *in-vitro* synthesised transcription factors as previously described. The proteins chosen for analysis were the Wnt pathway effectors Tcf3, Tcf7 and Lef1 and the early neuronal marker Sox2. We reasoned that the spatiotemporal expression profiles of these key developmental transcription factors coupled with the incidence of their binding sites within the 5' region of Motif3 made them suitable candidates. No interactions were observed between these proteins and Motif3 despite several attempts (data not shown). A series of experiments in which the binding of nuclear extract to Motif3 sequence was assessed did not provide satisfactory results, as the control reaction did not reproducibly

create a DNA-protein complex (data not shown). Taken together, these studies were not sufficient to advance our *in-vivo* evidence of Motif3 function.

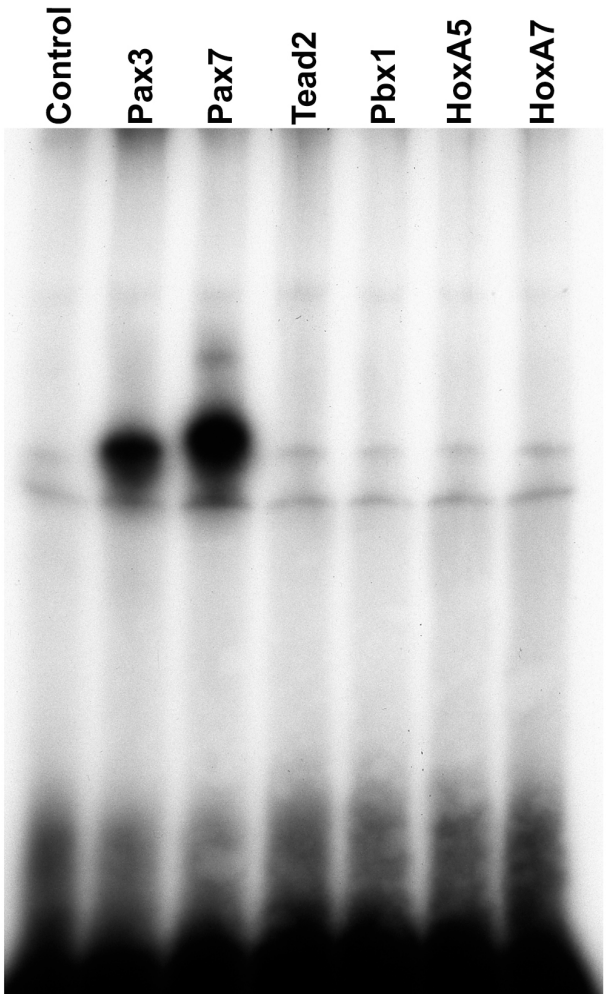


Fig. 7.11 – CNE1-Motif1 is bound by *in-vitro* synthesised Pax3 and Pax7

In-vitro synthesised Six2 protein does not bind the Motif1 DNA probe, as evident in the control lane. Addition of either Pax3 or Pax7 protein results in the formation of complexes that are both larger and more abundant than those observed in the control reaction. DNA-protein interactions are not observed in reactions including Tead2, Pbx1, HoxA5 or HoxA7 protein.

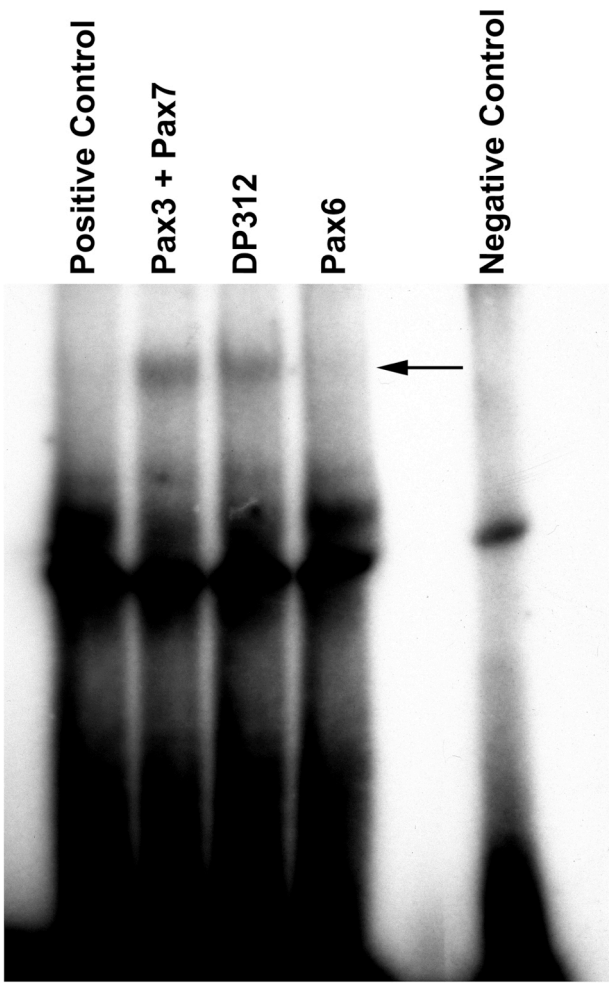


Fig. 7.12 – CNE1-Motif1 demonstrates specificity for Pax3/7 binding

A large DNA-protein complex is produced in the supershift control reaction, containing Motif1 DNA probe and nuclear extract prepared from chick spinal cord tissue. Addition of either Pax3 and Pax7 antibodies or DP312 produces a slower migrating complex, marked by the arrow within the autoradiogram. The inclusion of a Pax6 antibody does not produce a supershift when compared to the positive control lane, suggesting that this protein does not bind to the PD site represented by Motif1. The negative control reaction contains DNA probe in the absence of protein and demonstrates the non-specific bands within the autoradiogram.

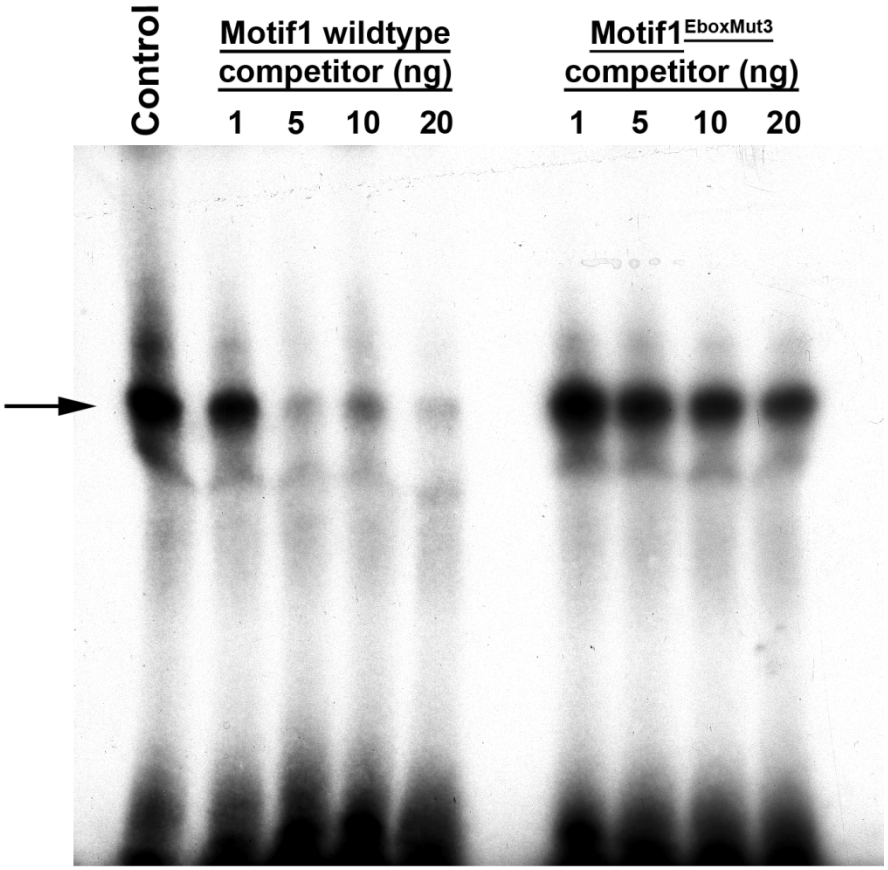


Fig. 7.13 – Mutation of the 3' region of Motif1 reduces its affinity for Pax7 binding

Titration of non-labelled Motif1 double stranded DNA probe to a 4-fold excess outcompetes the radiolabelled EMSA probe for Pax7 binding. This is visualised in the autoradiogram as a loss of complex formation following the point of equilibrium at 5 ng/reaction. In contrast, excess non-labelled Motif1^{EboxMut3} DNA cannot compete with the wild type sequence. These data indicate that this 2 bp substitution within the 3' region of Motif1 reduces the affinity of the PD site for Pax7 binding.

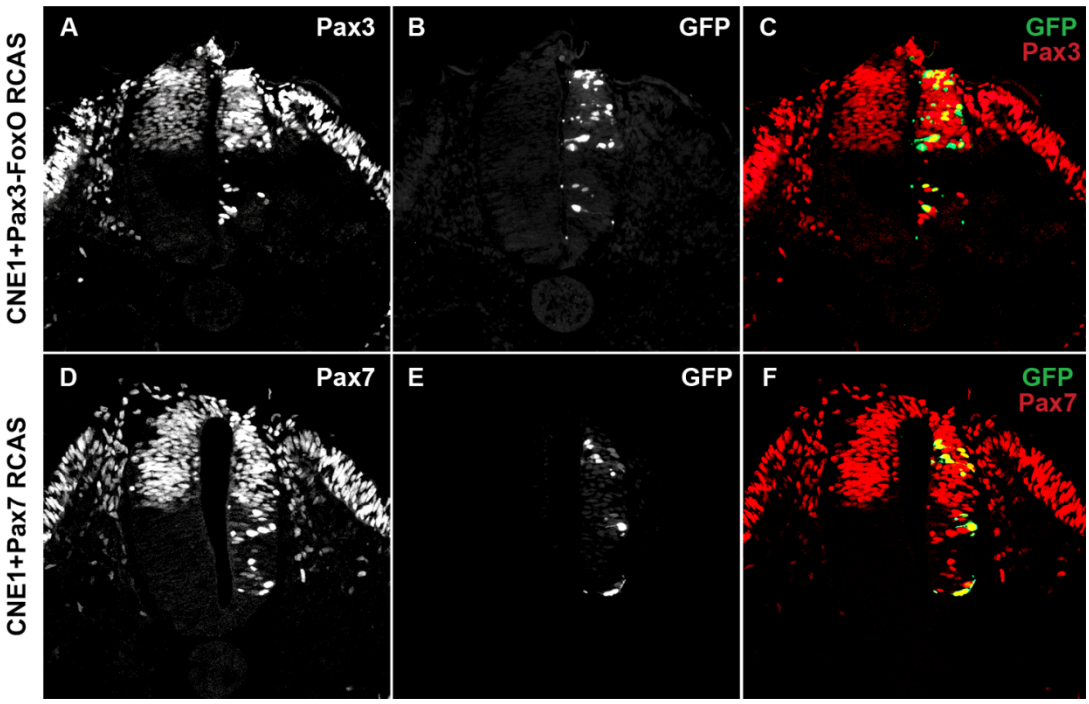


Fig. 7.14 – CNE1 mediates direct autoregulation and positive feedback

In-ovo electroporation of Pax3-FoxO RCAS is sufficient to induce both Pax3 protein expression and Citrine labelling in the ventral neural tube (A, B and C). Similarly, electroporation of Pax7 RCAS ectopically induces Pax7 protein and CNE1 activity (D, E and F).

7.4 – CNE1 is conserved between the zebrafish Pax3 paralogs

Functional dissection of CNE1 indicated that it contained a PD binding site that interacts with both Pax3 and Pax7 and that the integrity of this site was required for the correct activity of this regulatory element in spinal cord progenitors. We also demonstrated that CNE1 contained a putative HMG box binding site in Motif3 that appeared to be broadly required for its activity throughout the CNS. We aimed to extend our analysis of this enhancer to the zebrafish *pax3b* locus, in order to determine the conservation of these regulatory mechanisms following the whole genome duplication that occurred in the teleost lineage. Our previous studies were not able to identify any CNEs within the *pax3b* locus, as it was determined to be unsuitable for inclusion in comparative genomic alignments. Instead, we limited our search of the *pax3b* locus to intron 4, which was queried against all CNEs present in the Condor database by BLAST. This search identified a 38 bp region in the human genome that lies within CRCNE00001997, which we previously determined to be homologous to CNE1 (Greg Elgar, personal communication). This analysis also revealed that CNE1 was the only conserved region present within intron 4 of both paralogous zebrafish Pax3 loci, strengthening our hypothesis that this region performed a critical function in the transcriptional regulation of Pax3 expression.

Our initial analysis was sufficient to determine that the region of CNE1 homology identified by Condor BLAST was centred around the PD binding site of Motif1, which is 95% conserved between *pax3a* and *pax3b*. We expanded this alignment by adding 100 bp of 5' and 3' sequence from the *pax3b* locus to the BLAST hit and investigated the conservation of the remaining 3 motifs present in CNE1 using ClustalW2 (Fig. 7.15). These data indicated that the core of Motif2 was highly conserved, however *pax3b* exhibited a small 5' deletion and a larger 3' insertion within this region relative to *pax3a*. We were surprised to discover that Motif3 was not highly conserved in *pax3b*-CNE1 (40%), given the previously described requirement for this region in the canonical enhancer. Motif4 was determined to be 66.6% homologous, resulting from a 5' insertion and a 3' deletion in the *pax3b* locus. Taken together, these data indicated that CNE1 was poorly conserved in the *pax3b* locus, especially when compared to our previous analysis of this regulatory region in higher vertebrate genomes.

We reasoned that *pax3b*-CNE1 may retain some regulatory function *in-vivo* and that the differences in the architecture of this enhancer in the paralogous Pax3 loci

might account for the expression pattern of each transcript. A minimal 133 bp region of *pax3b* homology was defined using the location of each conserved motif, amplified from zebrafish genomic DNA using the primers listed in Table 2.1 and cloned into the PCR2.1 TOPO vector. Successful cloning was determined by sequencing before the insert was subcloned into the MiniTol2-Citrine vector by adding the desired restriction sites using the TOPO> MiniTol2 primer set listed in Table 2.1. Transgenic embryos injected with the defined minimal *pax3b*-CNE1 homology region and assessed at 24 hpf clearly demonstrated that this region maintained its regulatory potential (Fig. 7.16A). GFP staining was observed in progenitor cells within the dorsal spinal cord in all of these embryos (n=18). We frequently observed that transgene labelling was significantly reduced or absent in regions anterior to rhombomere 5, whereas *pax3a*-CNE1 was shown to be active throughout the mid and hindbrain. The specificity of CNE1 for CNS tissues was not maintained in the *pax3b* enhancer, with 66% (n=12/18) of embryos exhibiting strong transgene expression in muscle fibres. Transverse sections of trunk tissue at 24 hpf revealed that GFP staining was restricted to the Pax3/7 domain of the dorsal spinal cord (n=16) (Fig. 7.16B).

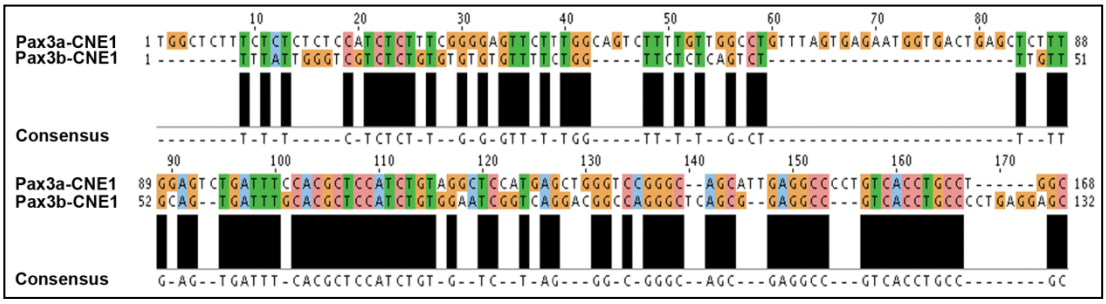


Fig. 7.15 – CNE1 is poorly conserved between the paralogous Pax3 loci

ClustalW2 alignment of sequence representing CNE1 within the *pax3a* and *pax3b* loci. Motif1 (positions 100-114) is 95% conserved. The core of Motif2 (positions 155-175) is homologous between the paralogs, however *pax3b* exhibits a small 5' deletion and a large 3' insertion in comparison to *pax3a*. Motif3 (positions 48-63) is poorly conserved, whereas Motif4 (positions 138-154) is partitioned into two blocks of homology.

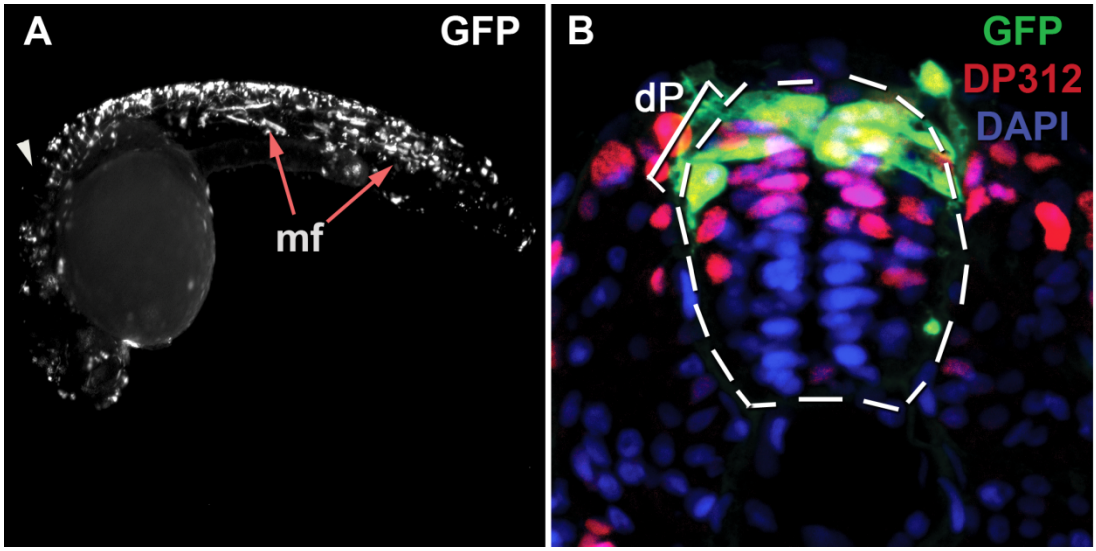


Fig. 7.16 – *Pax3b* CNE1 retains the primary function of the canonical enhancer

Despite poor overall sequence homology, *pax3b* CNE1 activity labels cells in the dorsal spinal cord (A) However, transgene expression does not appear to extend anteriorly beyond the hindbrain (arrowhead in A) and is frequently observed in muscle fibres (mf) (A). Cross sections at the level of the spinal cord reveals that GFP staining is restricted to the Pax3/7 expression domain of dorsal progenitors (dP).

Chapter 8: CNE3 defines the Pax3 expression domain in the neural plate

During the course of this investigation, we were surprised to discover that none of the mutations introduced within CNE1 resulted in an expansion of transgene expression within the ventral spinal cord. These data indicated that the activity of this enhancer was unlikely to be modulated by the binding of transcriptional repressors. However, our lineage experiments performed in mice suggested that Pax3 expression was downregulated in the intermediate regions of the spinal cord during the early stages of CNS patterning. In addition, we were unable to identify any molecular interactions that might implicate CNE1 in the induction of Pax3 transcription within the neural plate. Taken together, these data suggested that additional regulatory modules associated with the locus might act in concert with CNE1 to define and maintain the expression profile of Pax3 within the CNS.

8.1 – CNE3 possess a weak regulatory potential at 24 hpf

CNE3 is a 453 bp element located within the fourth intron of Pax3 that contains conserved homeobox and Tcf/Lef binding sites (Fig. 5.4). This region was identified as an interesting candidate regulatory region from our *in-silico* analysis, as many reports have demonstrated that the Wnt signaling pathway can both induce and modify Pax3 expression, discussed in Chapter 1.4.3. Nevertheless, our initial analysis of CNE3 activity *in-vivo* indicated that although CNE3 was active in the midbrain or hindbrain of 87% transgenic zebrafish assessed at 24 hpf, (n=20/23) (Fig. 8.1A), only very mosaic GFP staining was observed in the trunk. 78% of embryos exhibited Citrine labelling within the spinal cord, however upon closer inspection only 17% of these demonstrated GFP labelling within neural progenitors (n=3/18). Transverse sections of CNE3 transgenics confirmed that this region possessed a weak regulatory potential at 24 hpf, however it was possible to identify GFP staining within Pax3/7 expressing progenitors (Fig. 8.1B) (n=1/15).

Following our initial screen of CNEs associated with the Pax3 locus, the findings of a similar comparative genomics based study of Pax3 regulation was published (Degenhardt et al., 2010). The authors assessed the activity of a 967 bp conserved region, they termed Pax3-ECR2, and showed that this element was able to recapitulate the profile of *Pax3* expression throughout the CNS of mice at E10.5 and in zebrafish at 24 hpf (Degenhardt et al., 2010). We subsequently determined that CNE3 is

entirely included within the genomic interval represented by Pax3-ECR2 (Fig. 8.2). These data prompted us to reassess the activity of CNE3, however the results of these experiments were consistent with our previous observations. Accordingly, Citrine expression was observed within the anterior CNS but we were unable to reproducibly demonstrate that CNE3 was active within spinal cord progenitors at either 24 hpf (n=51) or 32 hpf (n=13).

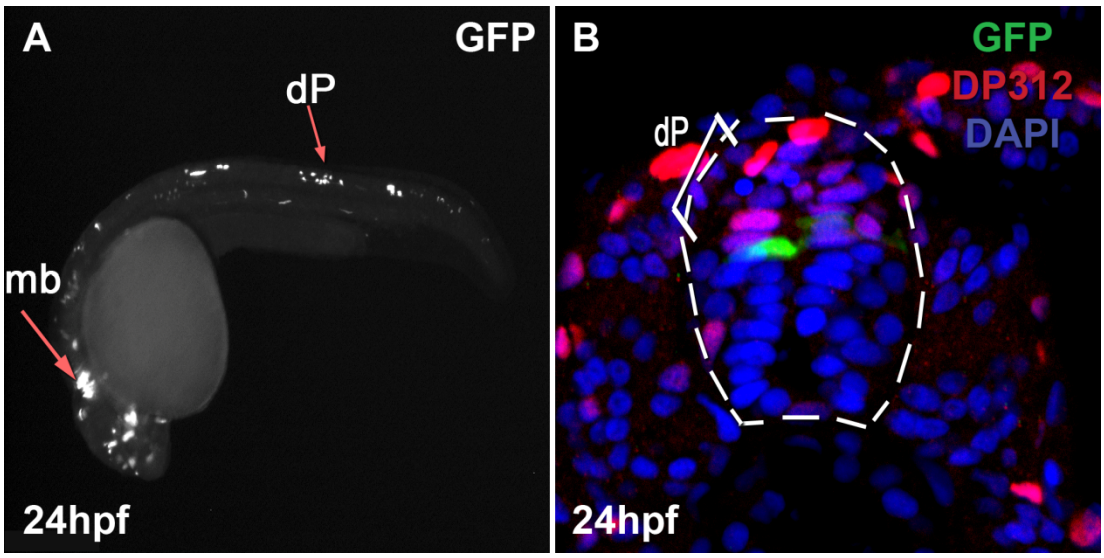


Fig. 8.1 – CNE3 is a weak enhancer at 24 hpf

CNE3 transient transgenic zebrafish assessed at 24 hpf reveal consistent GFP staining in the midbrain (mb), however dorsal progenitor cells (dP) within the neural tube are rarely labelled (A). Transverse sections of trunk tissue demonstrate that CNE3 mediated transcription is restricted to the Pax3/7 domain of dorsal progenitors (dP), when detected (B).

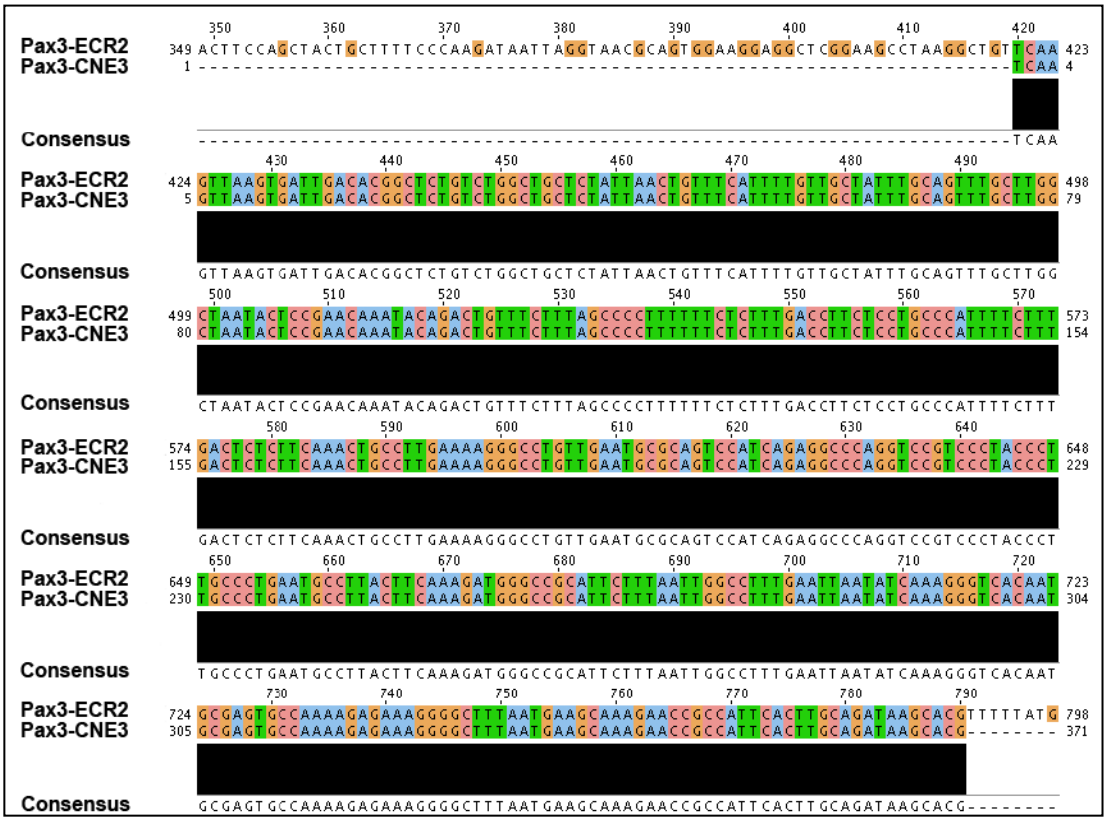


Fig. 8.2 – CNE3 is contained within Pax3-ECR2

ClustalW2 alignment of sequences representing Pax3-ECR2 and CNE3 in the mouse genome. These data demonstrate that CNE3 is homologous to the central portion of Pax3-ECR2.

8.2 – CNE3 is transiently active in the neural plate

The activity of Pax3-ECR2 was shown to be dependent upon the integrity of 5 Tcf/Lef sites, which when mutated in the context of a 403 bp minigene resulted in the loss of transgene expression in zebrafish embryos at 24 hpf (Degenhardt et al., 2010). These data suggested that ECR2 might facilitate the direct interaction of the Wnt signaling pathway with the Pax3 locus, providing a molecular mechanism of transcriptional induction within the CNS. Consistent with this hypothesis, ECR2 was shown to be active in the neural plate of stable transgenic mice at E8.5 (Degenhardt et al., 2010). We sought to investigate this further by using published data to validate the interaction of the Wnt pathway effectors with CNE3 *in-vivo*. Genome-wide binding of Tcf3 in mouse ESC revealed the direct interaction of this transcription factor with Pax3-ECR2 and furthermore, that the peak of sequenced reads overlapped with the location of CNE3 (Fig. 8.3) (Marson et al., 2008).

Taken together, these results were consistent with the conclusion that Pax3-ECR2 mediated the transcriptional output of the Wnt signaling pathway upon the Pax3 locus (Degenhardt et al., 2010). Furthermore, our observation that CNE3 was located beneath the peak of Tcf3 binding in ESC suggested that this region contained the key TFBS that mediated this interaction. We hypothesised that CNE3 may be temporally regulated during CNS development and that its dynamics may reflect the profile of Wnt signaling. To test this, we compared the activity of CNE3 with transgenic reporters of intracellular Wnt signaling during the first 24 hours of zebrafish development. Activated Wnt signaling was visualised using the *TCFsiam* transgenic line, which expresses eGFP under the control of a concatemer of Tcf binding sites derived from the promoter of the *Xenopus siamois* gene (Moro et al., 2012). We compared our data to published results documenting the activity of a second Wnt pathway reporter, *TOPdGFP,* which expresses destabilised GFP under the control of a 358 bp regulatory region containing 4 consensus Lef binding sites (Dorsky et al., 2002).

CNE3 transient transgenic embryos assessed at bud stage exhibited strong GFP staining within the posterior neural plate (n=10) (Fig. 8.4A). Analysis of the *TCFsiam* reporter demonstrated that the Wnt signaling pathway was activated in those cells marked by CNE3 mediated transcription (n=18) (Fig. 8.4E). These data were supported by observations of the *TOPdGFP* reporter, which exhibited GFP transcription throughout the posterior neural plate (Fig. 8.4I) (Dorsky et al., 2002). Transgene labelling by CNE3 was restricted within the posterior neural plate at the 6 somites stage

(n=7), as was GFP staining within *TCFsiam* embryos (n=12) (Figs. 8.4B and F). *TopdGFP* zebrafish assessed at this stage have been reported to exhibit a similar profile of Wnt pathway activation, however the authors also observed a domain GFP transcription within the midbrain hindbrain boundary that was not present in either CNE3 or *TCFsiam* embryos (Fig. 8.4J) (Dorsky et al., 2002). GFP staining was markedly increased within the midbrain, hindbrain and spinal cord of CNE3 transgenic embryos at the 18 somites stage (n=8) (Fig. 8.4C). *TCFsiam* embryos suggested that Wnt signaling active within the midbrain at this stage, however comparatively stronger GFP staining was observed within the posterior somites of the embryo (n=8) (Fig. 8.4G). The Wnt pathway has been demonstrated to be activated within the midbrain, hindbrain and anterior spinal cord of *TOPdGFP* zebrafish at the 18 somite stage, consistent with our observations of CNE3 labelling within these tissues (Fig. 8.4K) (Dorsky et al., 2002). By 24 hpf, GFP staining in CNE3 transient transgenic embryos was downregulated in the posterior CNS but maintained in the mid and hindbrain, as previously described (Fig. 8.4D). In contrast, *TCFsiam* embryos demonstrated that the Wnt pathway was active within the forebrain, the lateral line primordium and the posterior somites at this stage (n=9) (Fig. 8.4H).

Our analysis of the temporal dynamics of CNE3 mediated transcription revealed that transgene expression correlated with the activation of the Wnt signaling pathway in the neural plate. However during these experiments, we noted that the domain of cells that received Wnt signaling was not restricted along the medio-lateral axis of the early CNS and in the case of *TOPdGFP* embryos, was not limited to the neural plate (Dorsky et al., 2002). These results contrasted with the observed profiles of CNE3 activity and *pax3a* expression within the CNS, suggesting that the Wnt pathway alone does not provide sufficient positional information to establish the Pax3 expression domain. Given these data, we hypothesised that CNE3 also facilitated the interaction of repressors with the Pax3 locus and that a combination of positive and negative transcriptional inputs defined the domain of gene expression in the early embryo.

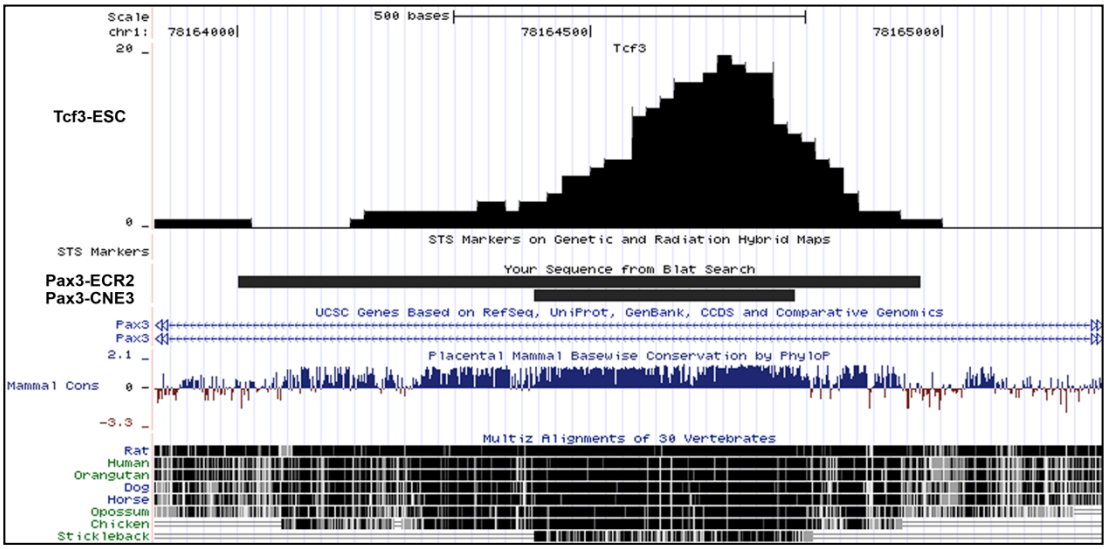


Fig. 8.3 – Tcf3 directly binds CNE3/ECR2 in mouse ESC

UCSC genome browser plot demonstrating the binding profile of Tcf3 in mouse ESC relative to both Pax3-ECR2 and CNE3. These data reveal that a peak of Tcf3 binding correlates with the position of CNE3 in the Pax3 locus, supporting the involvement of this enhancer in the Wnt mediated induction of Pax3 transcription.

ChIP-seq data published within Marson et al., 2008.

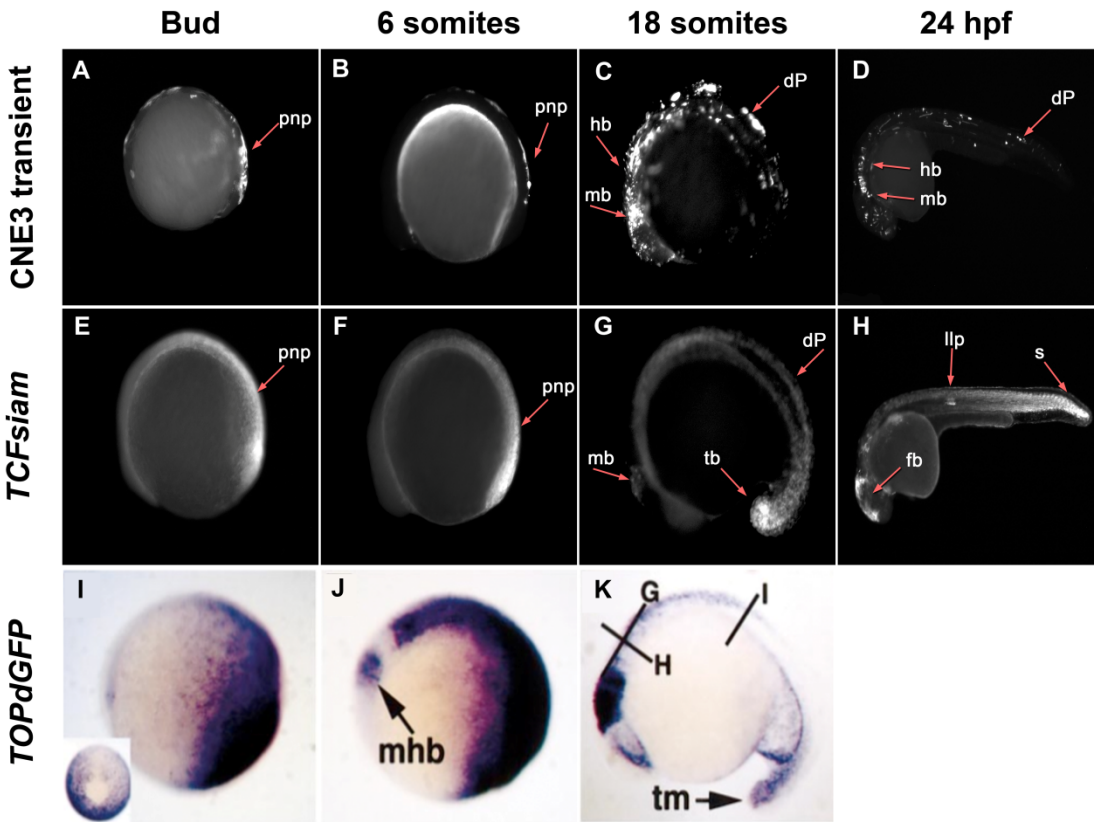


Fig. 8.4 – CNE3 labelling correlates with the temporal profile of Wnt signaling

GFP staining is restricted to the posterior neural plate (pnp) of embryos assessed at bud stage (A). Wnt pathway activation assessed in *TCFsiam* embryos is active within the posterior region of the neural plate whereas the *TOPdGFP* transgene is active throughout the posterior half of the embryo (E and I). CNE3 remains active in the posterior neural plate at the 6 somites stage (B), consistent with Wnt activity assessed in *TCFsiam* embryos (F). *TOPdGFP* zebrafish report active Wnt signaling throughout the posterior region of the embryo and within the midbrain at the 6 somites stage (J). CNE3 mediated transcription is increased in the midbrain (mb), hindbrain (hb) and dorsal progenitors (dP) within the anterior neural tube at 18 somites (C). *TCFsiam* embryos assessed at 18 somites report Wnt signaling in the tail bud (tb) posteriorly and weak transgene activation in the midbrain (G); however *TOPdGFP* transgenics exhibit robust pathway activation in the midbrain and hindbrain in addition to the tail bud. By 24 hpf, CNE3 activity is limited to the midbrain and single cells with the neural tube (D). By contrast, transgene expression in *TCFsiam* embryos is observed in the forebrain (fb), lateral line primordium (llp) and the posterior somites (s) (H). Taken together, these data demonstrate that CNE3 activity correlates with the temporal profile of Wnt pathway activation during CNS development.

Figures I, J and K were modified from Dorsky et al., 2002.

8.3 – Identifying putative repressor binding sites in CNE3

To advance this study, we employed comparative genomics to map the putative interactions of transcriptional repressors within CNE3. The 453 bp sequence of zebrafish CNE3 was used to identify the homologous regions in selected mammalian, amphibian, avian, reptilian and fish genomes by cross species BLAST searches, as previously described in Chapter 7.1. These experiments revealed that many vertebrate genomes contained a region of approximately 55 bp within CNE3 that did not share homology to the zebrafish enhancer, indicating that CNE3 was partitioned into two blocks of homology. The first region being approximately 57 bp and the second approximately 260 bp, we chose to select the larger homology block for further analysis and extracted this sequence from the desired genomes. This dataset was processed with the MEME suite of sequence analysis tools using the parameters described in Chapter 2.2.4. The results of this analysis revealed the presence of 5 highly conserved 15 bp motifs located within the 3' region of CNE3 (Table 8.1).

TomTom analysis of the PSPM representing Motif1 suggested this region contained a putative Sox binding site, whereas Motif2 matched the binding matrices of the transcriptional effectors of the retinoic acid signaling pathway (Table 8.1). We chose not to select these motifs for further analysis because they did not include putative repressor binding sites. Motif3 contained a consensus homeobox binding site that matched the binding matrices of several Nkx class transcription factors (Table 8.1). Members of this gene family are key fate determinants within the ventral spinal cord and have been shown to function as transcriptional repressors via recruitment of Groucho (Muhr et al., 2001). Our *in-silico* analysis of Motif4 indicated that it contained conserved binding sites for both Tcf3 and Tcf7, consistent with the interaction of this enhancer with the Wnt signaling pathway (Table 8.1). Motif5 also contained a Tcf3 and Tcf7 binding site in its 5' region, in addition to a 3' consensus homeobox binding site (Table 8.1). Taken together, these data suggested that Motif4 and Motif5 contained binding sites that facilitated the interaction of the Wnt signaling pathway with CNE3, whereas Motif3 and Motif5 contained binding sites for putative repressors of Pax3 transcription.

We chose to assess the requirement of Motif3 and Motif5 for the correct activity of CNE3 *in-vivo*. We reasoned that putative repressors would interact with CNE3 during the early stages of neural plate patterning to define the Pax3 expression domain, however we were aware that the zebrafish transient transgenic assay limited our ability

to interpret data collected at these stages. With these considerations in mind, we chose to assess the effect of motif deletions within CNE3 at both the 6 somites stage and at 24 hpf.

We did not observe any effect of Motif3 deletion at the 6 somites stage and Citrine expression was restricted to the posterior neural plate in all embryos assessed (n=17) (Fig. 8.5D). By contrast, 58% of Motif3 deletion transgenics exhibited Citrine labelling within spinal cord progenitors at 24 hpf, representing an increase in the activity of the enhancer relative to the wild type sequence (n=11/19). Furthermore, 82% of embryos expressing Citrine protein within neural progenitors demonstrated GFP staining in the ventral half of the spinal cord (n=9/11) (Fig. 8.5E and E'). We did not observe any defects in the distribution of Citrine protein in the neural plate of Motif5 deleted transgenic embryos assessed at the 6 somites stage (n=43) (Fig. 8.5F). However, GFP staining was observed in the spinal cord progenitors of 88% of embryos assessed at 24 hpf (n=28/32), of which 71% exhibited Citrine localisation in the ventral half of the tissue (n=20/28) (Fig. 8.5G and G'). These data are plotted and statistically analysed in Fig. 8.6.

Taken together, these data were consistent with a model in which the Wnt mediated activation of CNE3 is repressed in the medial neural plate or ventral neural tube to establish the domain of Pax3 expression. This was visualised in our results as a de-repression of Citrine protein expression in neural progenitors within the ventral spinal cord when either Motif3 or Motif5 were deleted from CNE3.

We next sought to determine the TFBS within these motifs that were required for the repression of CNE3 activity from the ventral neural tube. Our previous *in-silico* analysis of the phylogentically conserved binding sites within Motif3 indicated the presence of a canonical homeobox binding site and an overlapping Fox binding site (Table 8.1). We chose to induce mutations in both these sequences in the context of the wild type 453 bp enhancer by PCR using the primers listed in Table 2.4**.** Our analysis also identified a conserved homeobox binding site in Motif5, which was also mutated using the primer set listed in Table 2.4**.** We chose to assess the effects of these site mutations *in-vivo* at 24 hpf, as we were unable to define any significant changes in CNE3 activity at neural plate stages using the transient transgenic assay.

CNE3Motif3^{HoxMut} transgenic embryos exhibited a comparable de-repression of transgene activity to those in which the entire Motif was deleted. Accordingly, Citrine protein was observed in spinal cord progenitors of 49% of transgenic embryos (n=29/59). Of these, 83% exhibited GFP staining in ventrally located neuronal

progenitors (n=24/29 (Fig. 8.7C and C'). These data were supported by experiments in which the predicted Fox binding site within Motif3, which overlaps the Hox site, was mutated. 66% (n=29/44) of CNE3Motif3^{FoxMut} transgenics were observed to express Citrine protein within spinal progenitors and of these, 72% exhibited GFP staining within the ventral region of the tissue (n=21/29) (Fig. 8.7D and D'). 100% of CNE3Motif5^{HoxMut} embryos expressed the transgene within spinal cord progenitors (n=21/21), 76% of which exhibited a ventral expansion of GFP staining (n=16/21) (Fig. 8.7E and E'). Together, these data provide *in-vivo* evidence to suggest that the requirement of Motif3 and Motif5 for the repression of CNE3 activity was dependent upon the integrity of phylogenetically conserved Hox and Fox binding sites.

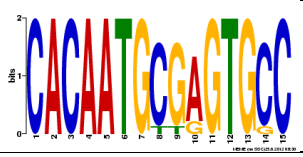
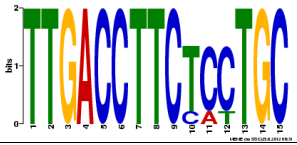
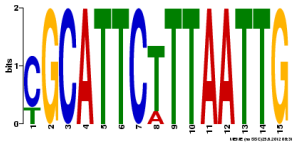
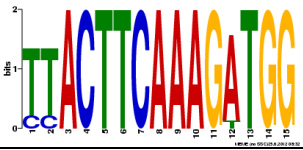
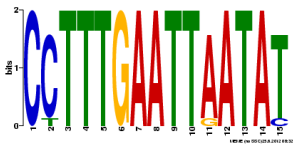
Name	Logo (12 vertebrate genomes)	Location within CNE3 (zebrafish)	Matching Matrices (JASPAR and UniPROBE)
Motif1		Bases 382-396	Gcm1, MATA1, Sox17, SOX10, Hbp1, Arid5a, h.
Motif2		Bases 206-220	NR4A2, Esrrb, RXR::RAR, RORA, Klf4, tll, Nr2f2, INO2, Zfp187, INO4, opa, TEC1
Motif3		Bases 343-357	Hmx, CG34031, CG11085, Hmx3, unc-4, B-H2, YHP1, Hmx2, NK7.1, B-H1, CG13424, CG15696, tup, hbn, slou, Hmx1, SPT2, Cart1, Uncx4.1, CG32532, Phox2b, Nkx2-5, bsh, Alx4, Lhx9, Ubx, En1, Nkx6-3, C15, ARID3A, Hoxc5, Otp, Prrx2, Hbp1, Dbx2, Hoxb4, repo, Dr, unpg, Nkx6-1, Hoxc4, Arx, Prop1, CG32105, Vsx1, Hoxa7, inv, Shox2, Rx, HGTX, Pou3f2, PHDP, Lim1, CG7056, Lhx5, Lbx2, OdsH, Pou3f4, Dfd, abd-A, Vsx2, TEC1, Esx1, Rhox6, al, Msx1, ftz, CG11294., CG4328, Lhx1, Pou3f1, Pou1f1, Hoxb7, Dll, Phox2a, Sox1, CG9876, cad, Scr, bap, Lhx3, Lmx1a, Elf3, Antp, HOXA5, YOX1, Foxa2, Lim3, Hoxa4, Pou4f3, Awh, btn, Nobox, Hoxc6, lab.
Motif4		Bases 326-340	Vnd, Bbx, Tcf3, YY1, Tcf7, pan, Gata1, TGA1A,
Motif5		Bases 359-373	CG7056, Lmx1a, Lmx1b, Arid3a, Hbp1, Prrx2, Lhx5, RME1, YOX1, Phox2b, Lhx1, Uncx4.1, Sox1, Hoxa7, Prop1, Otp, Hoxb4, Alx4, Lhx9, Hoxa4, Tcf3, Tcf7, Lhx3, Ubx, Hoxb8, Lim1, al, Cart1, unc-4, Glis2, Hoxc4, Nkx6-1, CG11294, Nkx6-3, Vsx1, Hmx, hbn, repo, Isl2, Hoxc6, Ptx1, Lef1, Pax7, CG32105, abd-A, Msx1, CG15696, Zfp187, CG34031, pan, Tlx2.

Table 8.1 – Annotation of conserved TFBS within CNE3

MEME analysis of 12 vertebrate CNE3 homologs reveals 5 statistically over-represented motifs. TomTom analysis of each motif PSPM demonstrates that Motif4 and Motif5 contain binding sites for Tcf/Lef transcription factors, whereas Motif3 and Motif5 include consensus sites for putative transcriptional repressors.

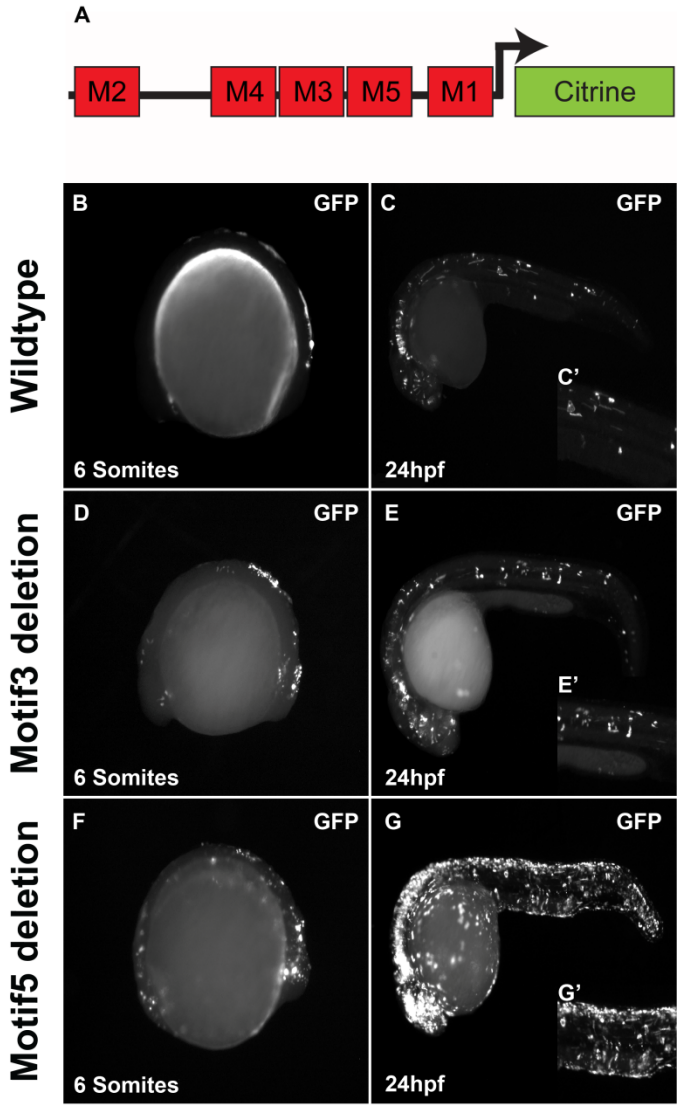


Fig. 8.5 – CNE3 activity is subject to transcriptional repression

Schematic highlighting the relationship between motifs identified within CNE3 (A). Motif3 deletion does not appear to affect the distribution of Citrine protein at the 6 somites stage (D), compared to wildtype sequence (B). By contrast, embryos assessed at 24 hpf exhibit an increase of CNE3 activity within the neural tube and labelled cells distributed across the DV axis of the tissue (E and E') which are not present in wildtype transgenics (C and C'). GFP staining also appears to be restricted to the posterior neural plate of Motif5 deletion transgenic embryos assessed at the 6 somites stage (F). However, robust Citrine protein expression is observed throughout the DV axis of the neural tube at 24 hpf (G and G').

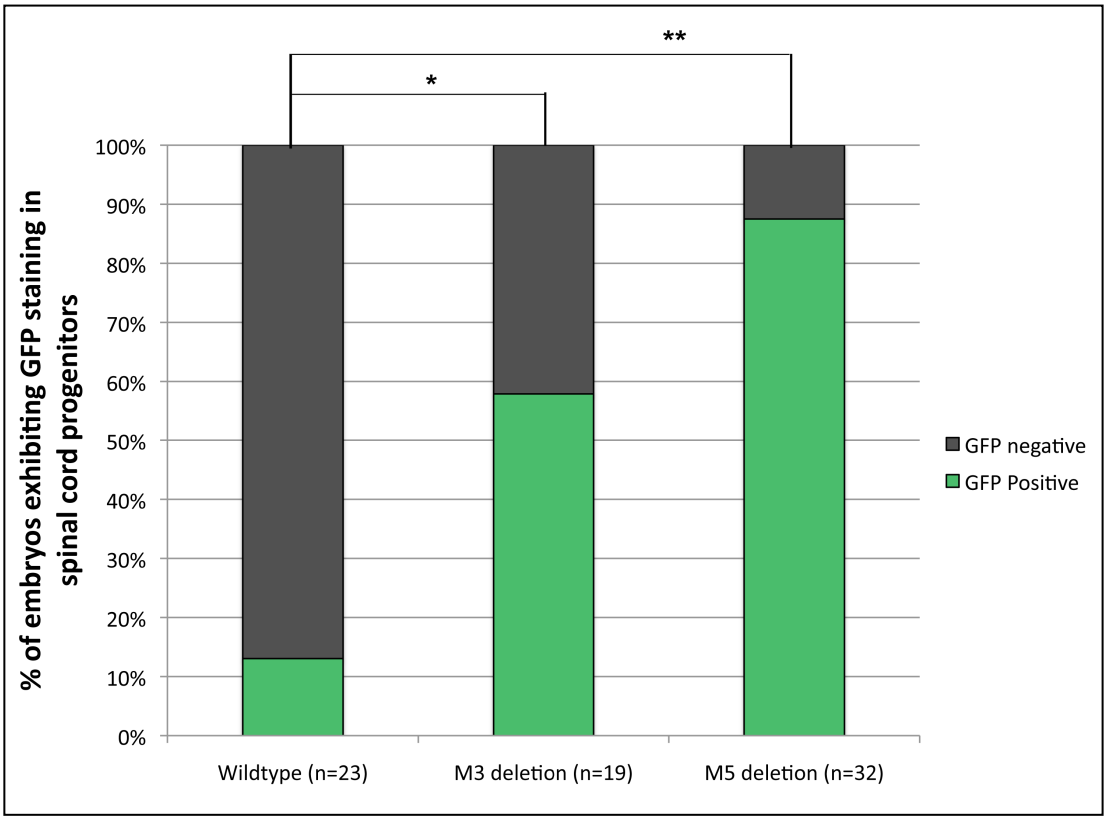


Fig. 8.6 – Quantification of the effects of motif deletions upon CNE3 activity

The percentages of transgenic zebrafish embryos exhibiting GFP staining in spinal cord progenitor cells (within the span of the yolk extension) assessed at 24 hpf is plotted. These data reveal that deletion of Motif3 results in a significant increase of CNE3 activity in the spinal cord. Similarly, loss of Motif5 increases the activity of CNE3 in spinal cord progenitors. These results are consistent with the hypothesis that Motif3 and Motif5 facilitate the interaction of putative transcriptional repressor with CNE3.

Two-tail Fisher's exact test (** = $P<0.001$ and * = $P<0.005$)

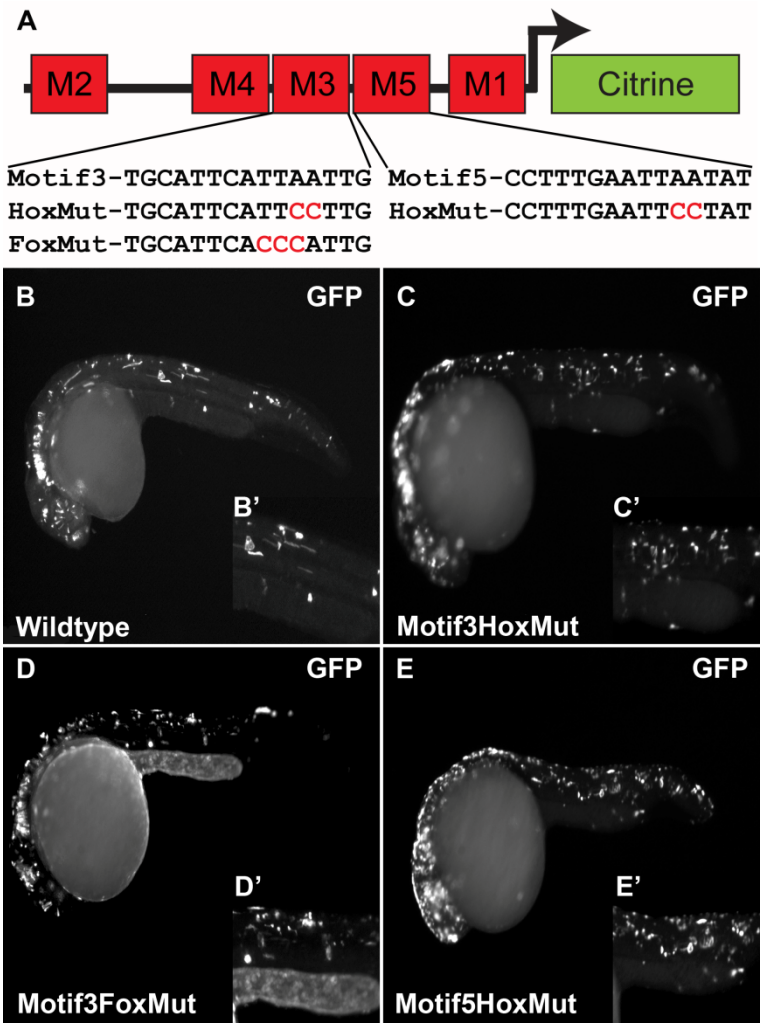


Fig. 8.7 – Conserved TFBS are required for CNE3 repression *in-vivo*

Schematic detailing the mutations iduced within conserved binding sites associated with Motif3 and Motif5 (A). Mutations induced within the overlapping Hox binding site within Motif3 results in a de-repression of CNE3 activity in the neural tube, compared to the wildtype sequence (B, B', C and C'). Similarly, mutation of the homeodomain binding site in Motif5 results in GFP staining within neural progenitors distributed across the DV axis of the tissue (E and E').

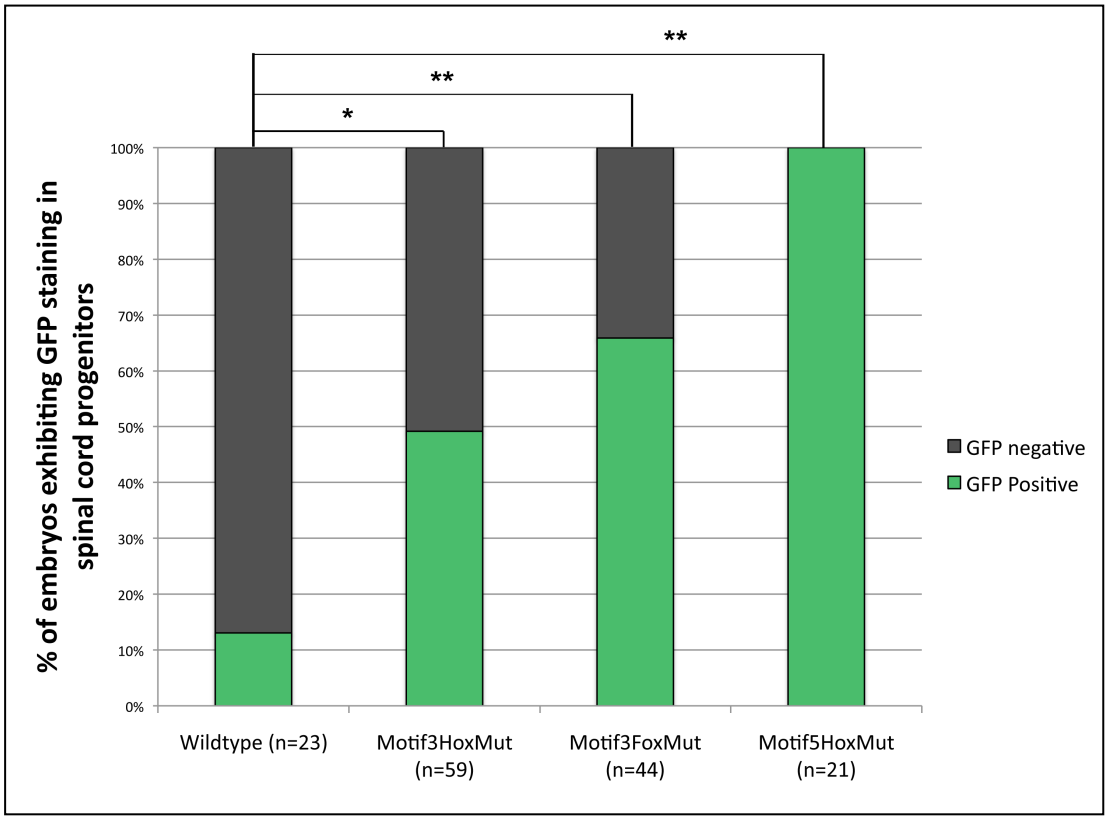


Fig. 8.8 – Quantification of the effects of site mutations upon CNE3 activity

The percentages of transgenic zebrafish embryos exhibiting GFP staining in spinal cord progenitor cells (within the span of the yolk extension) assessed at 24 hpf is plotted. These data indicate that mutation of the HD or Fox binding site within Motif3 results in a significant increase of CNE3 activity in spinal cord progenitors. In addition, mutation of the HD binding site within Motif5 results in a highly significant enhancement of CNE3 activity in the posterior neural tube. These data highlight the requirement of defined conserved TFBS for CNE3 repression *in-vivo*.

Two-tail Fisher's exact test (** = P<0.001 and * = P<0.005)

8.4 – CNE3 interacts with Nkx6.1 and Foxa2

Our *in-vivo* analysis of CNE3 regulation provided evidence to suggest that the activity of this regulatory region is subject to direct transcriptional repression, mediated via Motif3 and Motif5. We sought to investigate this further by determining the binding of candidate Pax3 repressors to CNE3 *in-vitro.* To this end, we designed a 48 bp EMSA probe that spanned the zebrafish genomic sequence containing Motifs 3, 4 and 5. The oligonucleotides and protocols used to create this DNA probe are described in Chapter 2.3 and Table 2.6. Binding of putative repressors was assessed in supershift EMSAs using total protein, obtained from chick nuclear extracts and antibodies raised against ventrally expressed transcription factors. All of the antibodies raised against putative repressors had been previously been demonstrated to recognise the antigen in chick tissue. We chose to include a β-catenin antibody in this screen to assess the interactions of these motifs with the Wnt signaling pathway, however previous experiments suggested that this antibody did not recognise the nuclear localised form of the protein in chick tissue (Ani Kicheva, personal communication).

The positive control reaction for this experiment, in which the CNE3Motif3-5 DNA probe was bound by total protein in the absence of antibodies, resulted in the formation of a reference complex (Fig. 8.9). Addition of β-catenin antibody to the reaction did not alter the migration of this complex, suggesting that either this protein did not interact with our probe or that the antibody employed was not suitable for its detection. Inclusion of an antibody against Nkx6.1 resulted in the supershift of the control complex and the stabilisation of a lower molecular weight complex. These data indicate that Nkx6.1 interacts with CNE3 Motif3-5 in this assay (Arrows in Fig. 8.9). No change in the migration of complexes was observed with antibodies against Nkx6.2, Pax6, Dbx1 or Dbx2 (Fig. 8.9). The addition of a Foxa2 antibody to the reaction resulted in the loss of all complexes from the autoradiogram, suggesting that complex formation was dependent on Foxa2 binding (Fig. 8.9). The negative control for this experiment, in which CNE3Motif3-5 probe was added to the reaction in the absence of protein, was used to determine the presence of non-specific labelling within the autoradiogram (Fig. 8.9).

The data indicated that both Nkx6.1 and Foxa2 proteins were able to bind the CNE3Motif3-5 probe *in-vitro,* in agreement with our previous *in-silico* annotation of putative binding sites for these factors within Motif3 and Motif5 (Table 8.1). We next aimed to assess the effect of the site mutations induced within these motifs upon the

binding of this putative repressor complex *in-vitro*. This was achieved by the creation of double stranded DNA probes from the oligonucleotides previously used to design mutations within CNE3, listed in Table 2.6. Binding of chick nuclear extract to the CNE3Motif3-5 probe resulted in the formation of a large DNA-protein complex in the positive control reaction for this experiment (Fig. 8.10). The binding of these proteins to CNE3 was impaired in experiments in which the homeobox or Fox binding sites in Motif3 were mutated (Fig. 8.10). A similar result was observed in reactions that contained a DNA probe in which the homeobox in Motif5 was mutated (Fig. 8.10). Together, these data suggest that the binding of a putative repressor complex to Motif3 and Motif5 within CNE3 is dependent upon the integrity of both the Hox/Fox binding site in Motif3 and the Hox binding site in Motif5, providing a molecular basis to our *in-vivo* observations.

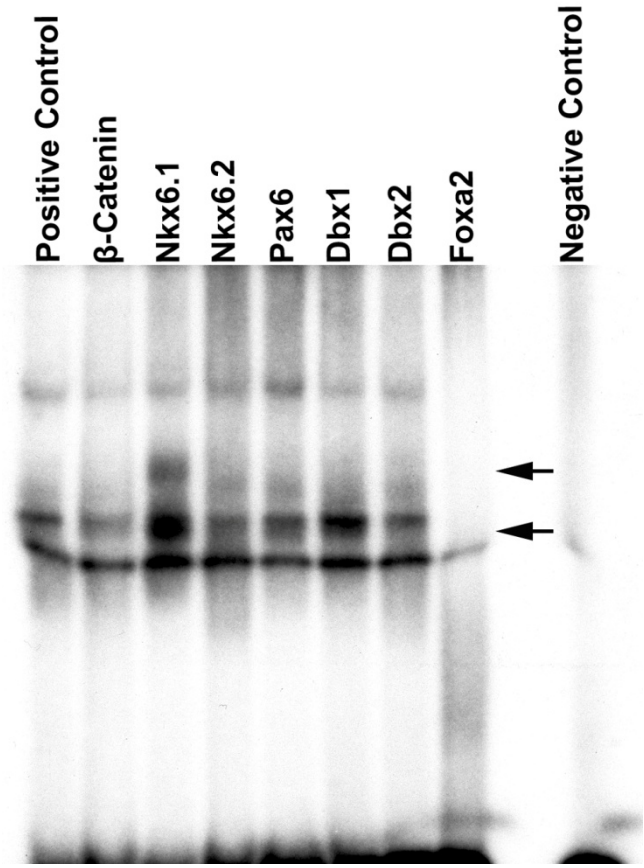


Fig. 8.9 – Nkx6.1 and Foxa2 bind CNE3 *in-vitro*

Supershift screen using chick nuclear extract and a DNA probe spanning Motif3-5 of CNE3. The positive control complex is not altered by the addition of antibodies raised against β-catenin, Nkx6.2, Pax6, Dbx1 or Dbx2. Inclusion of Nkx6.1 antibody in the EMSA reaction results in the formation of a slower migrating complex and the stabilisation of a lower molecular weight band, indicated by arrows within the autoradiogram. Addition of Foxa2 antibody disrupts the formation of the control complex, suggesting that the ability of the protein to bind DNA is impaired in the presence of the antibody. The negative control reaction contains DNA in the absence of protein and demonstrates the non-specific bands in the autoradiogram.

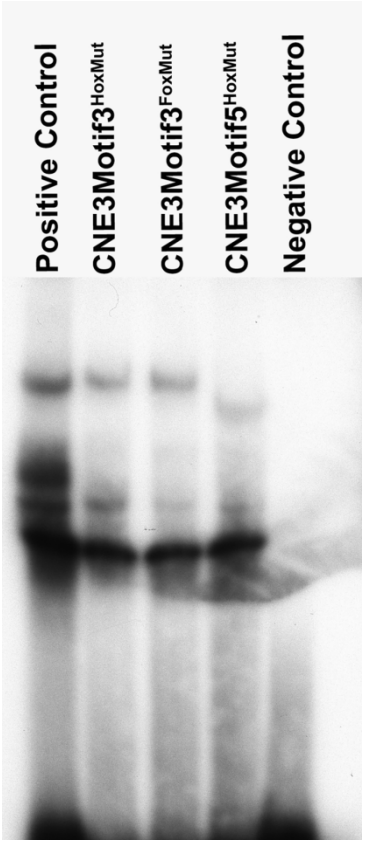


Fig. 8.10 – Mutations induced in CNE3 preclude the binding of putative repressors

CNE3Motif3-5 DNA forms a complex with total protein extracted from chick neural tube tissue. Mutation of either the homeodomain or Fox binding sites in the 3' region of Motif3 results in a loss of this control complex. Similarly, mutation of the homeodomain binding site within Motif5 precludes the binding of a putative repressor complex. The negative control reaction contains Motif3-5 DNA probe in the absence of nuclear extract.

Chapter 9: Discussion

Progenitor identity in the vertebrate neural tube is determined by the spatially organised expression profiles of transcription factors across the DV axis of the tissue. A series of selective cross-repressive interactions between these factors in the ventral neural tube form the basis of a GRN, which acts to define and maintain the boundaries between adjacent progenitor domains (Balaskas et al., 2012 ; Briscoe et al., 2000 ; Dessaud et al., 2008). By comparison, the regulatory interactions that act to define and maintain progenitor identity in the dorsal domains of the neural tube are less well understood. This study aimed to advance our understanding of the neural tube GRN by determining the regulatory logic that governs Pax3 expression during CNS development. Specifically, we sought to define the molecular mechanisms that induce Pax3 expression in the lateral regions of the neural plate and subsequently produce a sharp ventral boundary of gene expression at the level of the sulcus limitans.

These questions were addressed by firstly performing a lineage tracing study, which assessed the formation of the Pax3 domain in the neural tube and the fate of Pax3 expressing progenitors in the mature spinal cord. Our findings reveal that the position of the Pax3 expression domain is established before E9.5 in mice. Importantly, prior to E9.5, Pax3 is expressed in progenitors that later generate V0 interneurons, normally considered a ventral neuronal subtype. Our data also indicate that cell migration in the ventricular zone of the mouse neural tube is progressively restricted during development, consistent with published data obtained in chick embryos. The molecular basis of Pax3 expression was investigated in a comparative genomics study, which defined several novel tissue specific enhancers. Functional dissection of these regulatory regions provides evidence supporting the involvement of the Wnt signaling pathway in the induction of Pax3 transcription within the lateral neural plate. Furthermore, we discuss evidence establishing the direct input of repressors in this process. Finally, we demonstrate that Pax3 is subject to direct autoregulation throughout CNS development and positive feedback from Pax7 following neurulation. Together, these data support a model in which Pax3 expression in the neural tube is coordinated by the temporal activity of two distinct regulatory elements.

9.1 – The Pax3 domain is established by a switch in progenitor identity

Pax3 transcription is first reported at E8.5 of mouse development in the closing neural folds and the newly segregated dorsal neural tube (Goulding et al., 1991). Following the completion of neurulation, transcription is observed in all progenitors within the dorsal alar plate (Goulding et al., 1991). Given these data, we sought to use Pax3 as a molecular marker of the alar plate to test the hypothesis that this embryonic compartment is restricted to sensory fates in the functional spinal cord. Our findings demonstrate that Pax3 expression is not a molecular correlate of the mouse alar plate and provides novel insights into the cellular lineage of spinal interneurons and the regulatory mechanisms that establish the Pax3 domain during the early stages of CNS development.

Independent analyses of *Pax3*$^{Cre/+}$*/Rosa26* mice throughout the course of neural tube development revealed a domain of transgene expressing cells located adjacent to the Pax3 ventral boundary (Figs. 3.2, 3.3 and 3.5). This observation is reproducible, both along the AP axis and between individual embryos. The position of these cells within the neural tube is consistent with the expression profile of Dbx1, a molecular marker of the p0 progentor domain (Pierani et al., 2001). In agreement with this conclusion, this domain of Pax3 progeny represents Evx1 expressing progenitors fated to give rise to V0 interneurons in embryos assessed at E11.75 (Fig. 3.6) (Moran-Rivard et al., 2001). These data indicate that progenitors of this neuronal subtype transiently express Pax3 within the neural plate before being respecified to a more ventral identity. Despite being derived from Pax3 expressing progenitors, V0 interneurons do not appear to require Pax3 for their development, as Evx1 expression is maintained in *Pax3*$^{-/-}$*;Pax7*$^{-/-}$ mouse embryos (Mansouri and Gruss, 1998). Thus, our findings provide novel insights into the lineage of V0 interneurons and indicate that this ventral neuronal subtype is derived from cells that have transiently expressed Pax3 early in development.

Our lineage trace data demonstrate that Pax3 expression is consistently induced in the lateral and intermediate regions of the mouse neural plate (Figs. 3.2, 3.3 and 3.5). Furthermore, our findings demonstrate that isolated clones of Pax3 progeny are frequently observed in the ventral domains of the neural tube (Fig. 3.4). These data indicate that Pax3 expression in mice is induced in the intermediate neural plate before becoming restricted to the alar plate, in contrast to the results of previous expression studies (Goulding et al., 1991). Studies performed in chick embryos support this conclusion by revealing that Pax3 transcription is induced throughout the open neural

plate and subsequently limited to the neural folds prior to neurulation (Bang et al., 1997 ; Goulding et al., 1993 ; Liem et al., 1995). Furthermore, our expression studies performed in zebrafish indicate that *pax3a* transcription is induced in the intermediate regions of the posterior neural plate at bud stage and only becomes restricted to the lateral limits of the tissue at the 6 somite stage (Fig. 4.1). Together, these studies provide data to support a model in which Pax3 expression in the open neural plate is regulated by transcriptional repression.

Our findings indicate that progenitor identity is dynamically reassigned during the early stages of neural plate patterning. These data are in agreement with the results of investigations assessing the acquisition of progenitor identity in the ventral domains of the neural tube. Genetic tracing of the p0 marker Dbx1 in mice reveals extensive transgene labelling of cells both within the endogenous expression domain and in the more ventral p1, p2 and pMN (Dessaud et al., 2010). Similarly, all Nkx2.2 expressing p3 progenitors transit through a stage in which they express the pMN marker Olig2 (Chen et al., 2011 ; Dessaud et al., 2007). These data are consistent with a model in which basal plate progenitors are progressively ventralised by increasing levels of intracellular Shh signaling over time (Briscoe et al., 2000). Taken together, the results of these lineage studies demonstrate that progenitor identity is both plastic and dynamic at early stages of development across the DV axis of the neuroepithelium. Moreover, these data highlight the critical functions that transcriptional networks perform in order to ensure the maintenance of cell identity at later stages of development.

9.2 – Progenitor dispersal is restricted during mouse CNS development

Previous studies assessing clonal dispersion in the chick neuroepithelium have demonstrated that cell migration in the DV axis becomes restricted following tissue patterning (Erskine et al., 1998 ; Leber and Sanes, 1995). Our findings support these conclusions by demonstrating extensive cell migration across the medio-lateral axis of the mouse neural plate before E9.5. However, independent analyses of the Pax3 lineage demonstrate that once established, the position of the Pax3 expression domain remains constant throughout the later stages of neural tube patterning and neurogenesis. Thus, cell migration in the DV axis of the mouse neuroepithelium exhibits a similar temporal restriction to that observed in chick embryos.

The extent to which cells migrate within the DV axis of the neural tube is broadly correlated with the species-specific expression of cell adhesion molecules at the

presumptive sulcus limitans, discussed in Chapter 1.3. In comparison to these studies, the temporal expression profiles of cadherin family members during the development of the mouse neural tube are poorly described. Reports have shown that *Cadherin7* is not expressed in the neural tube at any point during embryogenesis (Moore et al., 2004). However the mouse *F-cadherin* homolog, *Cadherin20,* is reported to be expressed throughout the ventral domains of the neural tube at E10.5 (Moore et al., 2004). By contrast, *cadherin7* is expressed throughout the alar plate of the rat hindbrain whereas *cadherin20* appears to be expressed in a complementary profile within the basal plate (Takahashi and Osumi, 2008). These data raise questions concerning the molecular basis of cadherin transcription in the neural tube and the apparent lack of conserved regulatory mechanisms between species. Studies in the chick neural tube provide evidence demonstrating that *Cadherin-7* expression is positively regulated by low levels of Shh signaling and repressed by ectopic Pax7 *in-vivo* (Luo et al., 2006 ; Stamataki et al., 2005). Moreover, functional dissection of a conserved Cadherin7 enhancer suggests that these regulatory interactions may be direct (Prasad and Paulson, 2011). These studies provide evidence to suggest that, at least in chick embryos, the establishment of the Pax3/7 domain might directly affect progenitor cell migration in the neural tube by positioning *Cadherin-7* expression at the sulcus limitans.

9.3 – Conservation of non-coding sequence is predictive of functional enhancers

Previous studies investigating the regulatory logic of Pax3 expression in the neural tube have focused on a 1.6 kb CRM located proximal to the promoter in the mouse genome (Degenhart et al., 2010 ; Millewski et al., 2004 ; Natoli et al., 1997 ; Pruitt et al., 2004). By contrast, this investigation aimed to define the CRMs associated with the Pax3 locus *de-novo* using a stringent implementation of comparative genomics. Our alignments clearly demonstrate that the 1.6 kb upstream element is not conserved across the evolutionary span of vertebrates, suggesting that this CRM arose as a secondary Pax3 enhancer during mammalian evolution (Fig. 5.1). This conclusion is supported by a report demonstrating that deletion of this CRM from the mouse genome does not affect gene expression (Degenhardt et al., 2010).

Our analysis of the Pax3 locus revealed 6 CNEs located within the large 4th intron of the gene, all of which correlate with regions previously identified by similar bioinformatics methods (Table 5.1) (Woolfe et al., 2005 ; Woolfe et al., 2007). We provide evidence demonstrating that 5 of these regions act as tissue specific enhancers

in the zebrafish embryo, emphasising the ability of comparative genomics to predict functional CRMs. Our alignments identified 4 CNEs that are sufficient to activate transcription within regions of the embryo that endogenously express Pax3. We also define CNE5 as an enhancer that is predominantly active in cells that do not express Pax3 at any point of their development. This finding raises interesting questions concerning the function of this CRM during embryogenesis and the possibility that despite its location, it may function independently of the Pax3 locus. Thus, our functional assessment of Pax3 CNEs indicates that the complex spatiotemporal profile of gene expression is produced by the combined activity of several tissue specific enhancers.

Advances in post-genomic technologies have facilitated the identification of chromatin signatures that both define putative enhancers and predict their temporal activation, discussed in Chapter 1.2. We sought to determine the location of our Pax3 CNEs relative to the distribution of H3K4Me3 and H3K27ac marks to investigate the correlation between putative enhancers identified by these two methods. This analysis was facilitated by a study that documented the profile of selected histone modifications across the zebrafish genome during the first 48 hours of development (Bogdanovic et al., 2012). Comparison of these two datasets reveals that all Pax3 CNEs are located in regions marked by both H3K4Me3 and H3K27ac at some point during development, consistent with these regions being functional enhancers and residing in areas of active chromatin (Fig. 9.1). This analysis builds upon the previous observation that 90% of all p300 peaks in the mouse genome overlap with areas of conserved sequence (Visel et al., 2009). Together, these data reveal the predictive power of both comparative and post-genomic methods for enhancer discovery and highlight the benefit of employing both techniques.

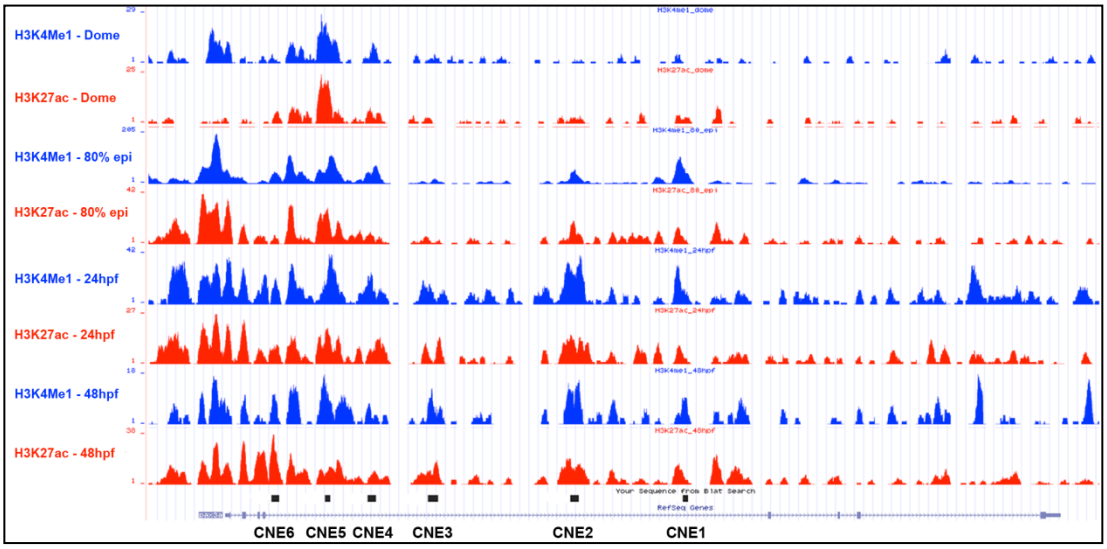


Fig. 9.1 –Pax3 CNEs correlate with the chromatin signatures of active enhancers

UCSC genome browser plot of H3K4Me1 and H3K27ac ChIP-seq profiles derived from zebrafish embryos at 4 stages of development. The location of each CNE identified by comparative genomics is marked in the plot. These data reveal that each CNE is labelled by both H3K4Me1 and H3K27ac at some point during the first 48 hours of development. ChIP-seq data published within Bogdanovic et al., 2012.

9.4 – CNE3 mediates the induction of Pax3 transcription

Several studies have described the effect of Wnt pathway manipulation on the expression of Pax3 within the neural tube, discussed in Chapter 1.3. The molecular basis of this interaction was established by the identification of Pax3-ECR2, a 967 bp conserved enhancer located within the 4th intron of the gene. ECR2 activity appears to recapitulate Pax3 expression throughout the course of CNS development in mice (Degenhardt et al., 2010). Furthermore, the transcriptional activity of ECR2 is dependent on the integrity of 5 Tcf/Lef binding sites (Degenhardt et al., 2010). Pax3-ECR2 is represented in our dataset as CNE3, a 453 bp element that possesses a weak regulatory potential in zebrafish embryos assessed at 24 hpf. We build upon the results of previous studies by demonstrating that the Wnt pathway effector, Tcf3, is bound to the region represented by both ECR2 and CNE3 in mouse ESCs (Fig. 8.3) (Marson et al., 2008). Furthermore, we reveal that the spatiotemporal activity of CNE3 correlates with that of the Wnt signaling pathway during zebrafish development (Fig. 8.4). Thus, our data provide support for a model in which Pax3 transcription is initiated in the neural plate by the direct binding of Wnt pathway effectors to ECR2/CNE3.

Strikingly, analysis of transgenic zebrafish embryos revealed that the profile of Wnt signaling in the neural plate is not patterned across the medio-lateral axis of the tissue (Fig. 8.4). These data are consistent with reports of Wnt activity during the early stages of mouse development (Maretto et al., 2003). By contrast, our expression studies in zebrafish (Fig. 4.1) and lineage studies in mice (Fig. 3.2) demonstrate that Pax3 transcription is excluded from progenitors in the medial neural plate. These observations indicate that Wnt signaling alone does not provide sufficient positional information to establish the Pax3 expression domain. Our functional dissection of CNE3 revealed that Motif3 and Motif5 are required for the repression of transgene activity from medial progenitors *in-vivo* (Fig. 8.5). Bioinformatic analyses of these sequences indicated the presence of conserved homeobox and Fox binding sites, which when mutated results in the de-repression of CNE3 activity *in-vivo* (Table 8.1 and Fig. 8.6). These data are supported by the observation that both Nkx6.1 and Foxa2 can bind the sequence spanning these motifs in EMSA experiments (Fig. 8.7).

Despite their ability to interact with CNE3 *in-vitro*, the *in-vivo* function of Nkx6.1 and Foxa2 in the context of Pax3 regulation requires further investigation. Nkx6.1 is among the first transcription factors expressed in the medial neural plate of zebrafish, *Xenopus* and chick embryos, consistent with the hypothesis that it might

function to repress Pax3 (Cheesman et al., 2004 ; Dichmann and Harland, 2010 ; Qiu et al., 1998). However, the observation that ectopic Nkx6.1 expression in the *Xenopus* neural plate does not repress Pax3 transcription suggests that our *in-vitro* data does not accurately reflect the *in-vivo* relationship between these proteins (Dichmann and Harland, 2010). Furthermore, *Pax7* expression in the neural tube is not de-repressed in both *Nkx6.1*$^{-/-}$ and *Nkx6.1*$^{-/-}$*;Nkx6.2*$^{-/-}$ mutants (Sander et al., 2000 ; Vallstedt et al., 2001). One would predict that de-repression of Pax3 would result in the ventral expansion of Pax7, given the patterning phenotype in mice expressing Pax3 under the control of the *Hoxb-4* enhancer A region (Tremblay et al., 1996). Similarly, Foxa2 is expressed in presumptive FP cells at the midline of the neural plate and throughout the developing endoderm (Ruiz i Altaba et al., 1995 ; Sasaki and Hogan, 1994 ; Strahle et al., 1993). A recent ChIP-seq study identified a peak of Foxa2 binding that overlaps with both ECR2 and CNE3 in ESCs differentiated to a midbrain dopaminergic identity (Metzakopian et al., 2012). However the same report demonstrated that, at least in the midbrain, loss of both *Foxa1* and *Foxa2* does not result in a de-repression of *Pax3* transcription (Metzakopian et al., 2012).

Taken together, our analysis of CNE3 supports the role of this enhancer in the Wnt mediated induction of Pax3 transcription within the neural plate. In addition, we establish a role for spatially expressed repressors in this process by defining regulatory motifs within CNE3 that are required for the restriction of transgene expression to progenitors within the lateral neural plate. Furthermore, we demonstrate that conserved homeobox and Fox binding sites within these motifs mediate the binding of putative repressors to CNE3. However, the identity of the transcription factors that bind these sites and their *in-vivo* activity in the context of Pax3 regulation requires further investigation.

9.5 – Pax3 transcription is maintained by CNE1 at later stages of development

Our analysis of CNE3 demonstrates that this enhancer possesses a weak regulatory potential at 24 hpf in zebrafish embryos and rarely labels spinal cord progenitors (Fig.8.1). Studies of *TCFsiam* transgenic zebrafish reveal that progenitor cells within the neural tube are not subject to active Wnt signaling from the 18 somites stage, a conclusion supported by reports of *TOPdGFP* transgenics (Fig.8.4) (Dorsky et al., 2002). Furthermore, analysis of BAT-Gal transgenic mice indicates that the Wnt pathway is not transcriptionally active in the anterior regions of the neural tube at E10.5

(Maretto et al., 2003). These studies demonstrate that the maxima of Pax3 expression in spinal cord progenitors occurs in the absence of Wnt pathway activation. Moreover, many reports have shown that the ventral limit of Pax3 expression does not correlate with the boundary of candidate homeodomain or Fox repressors in the neural tube (Dessaud et al., 2008 ; Jessell, 2000 ; Lewis, 2006). Together, these findings indicate that that the activity of CNE3 is transient and does not contribute the maintenance of Pax3 expression at later stages of development.

Our study identified a second CNS specific enhancer associated with the Pax3 locus, termed CNE1. We demonstrate that this regulatory region is transcriptionally active in the intermediate and lateral regions of the open neural plate and subsequently recapitulates Pax3 expression throughout CNS development (Figs. 6.4 and 6.5). This *in-vivo* data is supported by the observation that CNE1 is marked by H3K4Me1 in the presence of H3K27ac at the 80% epiboly stage of zebrafish development, a chromatin signature indicative of transcriptionally active enhancers (Fig. 9.1) (Bogdanovic et al., 2012 ; Creyghton et al., 2010 ; Rada-Iglesias et al., 2011). Analysis of p300 binding across the Pax3 locus at E11.5 demonstrates that CNE1 is the only active enhancer associated with the Pax3 locus and furthermore, that this interaction is specific to CNS derived tissues (Fig. 6.3) (Visel et al., 2009). By contrast CNE3 is not bound by p300 in any ChIP-seq dataset derived from E11.5 mouse tissue (Visel et al., 2009). Together, these data indicate that CNE1 regulates Pax3 transcription throughout development, whereas CNE3 is not transcriptionally active at later stages.

Functional dissection of CNE1 reveals that the activity of this region *in-vivo* is dependent on the presence and integrity of Motif1 (Figs. 7.2 and 7.5). We utilise a combination of *in-silico* sequence analysis and *in-vitro* binding data to functionally annotate Motif1 as a 15 bp PD binding site that appears to be preferentially occupied by group III Pax genes (Figs. 7.3, 7.7 and 7.8). The observation that both Pax3/7 are capable of binding the same DNA sequence is consistent with the fact that the PDs encoded by these genes share 93% homology at the amino acid level. The specificity of this site for Pax3/7 binding *in-vitro* can be attributed to the divergent PDs encoded by each class of Pax gene. This is exemplified by comparisons of the mouse Pax3 PD with that encoded by group II and group IV genes, revealing a maximum of 73% and 66% homology at the amino acid level, respectively.

Previous *in-vitro* selex based studies have provided definitions of the consensus site bound by an isolated Pax3 PD fusion protein (Chalepakis and Gruss, 1995 ; Epstein et al., 1996). Comparison of these published sites with Motif1 reveals significant

homology to the 5' portion of the PD binding site, however no consensus sequence is produced in the 3' region (Fig. 9.2). This result is consistent with our previous sequence analysis of published Pax binding sites that originally identified Motif1 as a putative PD site (Fig. 7.3). Our mutagenesis series within Motif1 suggests that RBPJ mutation induced within the 5' region of the PD site does not disrupt CNE1 function *in-vivo* (Fig. 7.4). Furthermore, mutations induced in the 3' region of the PD site both decrease CNE1 transcriptional activity *in-vivo* and Pax7 binding *in-vitro* (Figs. 7.5 and 7.9). These data suggest that *in-vitro* selex favours the identification of a 5' half site that is recognised by the PDs of all Pax genes, whereas Motif1 represents a more accurate description of a complete Pax3/7 PD binding site.

Expression of an obligate transcriptional activator form of Pax3, Pax3-FoxO, throughout the chick neural tube results in the ectopic activation of CNE1; establishing the molecular basis of direct Pax3 autoregulation during CNS development (Fig. 7.10). Previous reports have provided evidence to suggest that positive autoregulation may be a common property of Pax3 transcription during development. P19 EC cells cultured in the presence of dimethyl sulfoxide give rise to muscle fibres that express Pax3 during their progenitor phase (Ridgeway and Skerjanc, 2001). Stable expression of Pax3 fused to the repressor domain of Engrailed results in the repression of endogenous Pax3 transcription during this process, assessed by Northern blot (Ridgeway and Skerjanc, 2001). In a similar experiment, electroporation of a Pax3–Engrailed fusion protein in the ophthalmic trigeminal placode was shown to extinguish endogenous Pax3 expression (Dude et al., 2009). These studies provide evidence supporting the requirement of Pax3 autoregulation during development. This investigation builds upon published data by demonstrating that Pax3 transcription is subject to direct autoregulation throughout the process of CNS development and provides the molecular mechanism by which this is achieved.

Our findings also demonstrate that ectopic Pax7 expression *in-ovo* induces CNE1 activity in the ventral neural tube, indicating that positive feedback plays a role in the regulation of Pax3 transcription (Fig. 7.10). This conclusion is supported studies demonstrating that ectopic Pax7 expression induces Pax3 transcription cell autonomously within the dorsal midbrain (Agoston et al., 2012 ; Matsunaga et al., 2001). Pax7 expression in mice and chick embryos is induced in the neural tube as it closes (Jostes et al., 1990 ; Kawakami et al., 1997). These data are complemented by our analysis of *pax7a* and *pax7b* transcription during zebrafish development (Figs. 4.2

and 4.4). These results indicate that the Pax7 mediated positive feedback can only regulate Pax3 transcription following neural tube formation.

Taken together, our analysis of CNE1 establishes this region as a novel, CNS specific enhancer of Pax3. We demonstrate that CNE1 is transcriptionally active within the lateral neural plate of zebrafish embryos and subsequently recapitulates Pax3 expression during CNS development. We determine that the transcriptional activity of CNE1 in neural tube progenitors is dependent on the presence of two highly conserved regions, named Motif1 and Motif3. We demonstrate that Motif1 represents a novel Pax3/7 specific PD binding site that mediates direct autoregulation throughout CNS development and positive feedback from Pax7 following neurulation.

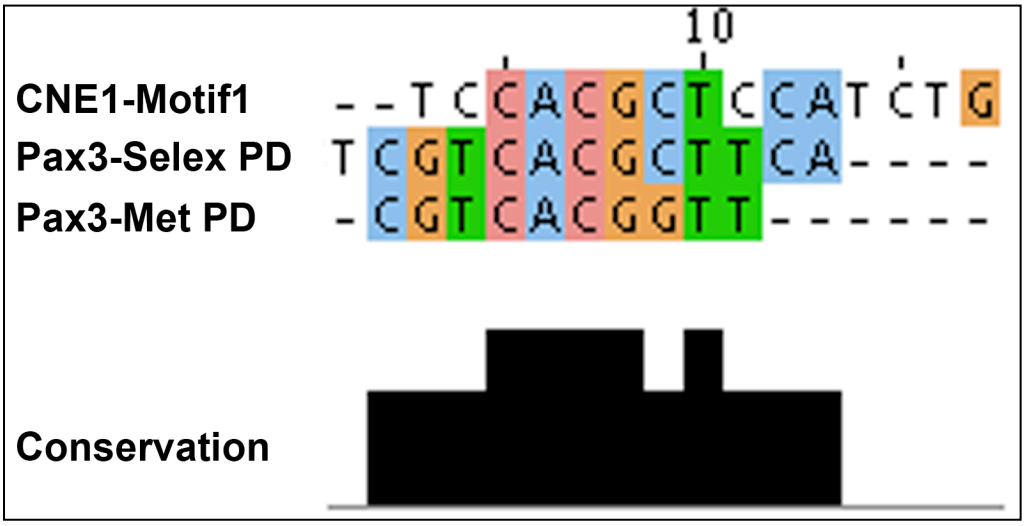


Fig. 9.2 – CNE1-Motif1 represents a more complete definition of the Pax3/7 PD binding site

ClustalW2 alignment of CNE1-Motif1 with *in-vitro* selex derived Pax3 PD consensus binding sites. The 5' region of each input sequence shares a high degree of homology, whereas CNE1-Motif1 is contains 4 additional nucleotides at the 3' terminal. Our data demonstrates that mutation of this 3' region of CNE1-Motif1 disrupts enhancer activity *in-vivo* and reduces Pax7 binding *in-vitro*. Consequently, CNE1-Motif1 represents a more accurate definition of a complete Pax3/7 binding site.

9.6 – *Pax3b* expression is dependent on autoregulation and positive feedback

The induction of *pax3b* expression in the posterior neural plate of zebrafish embryos is delayed in comparison to *pax3a* (Minchin and Hughes, 2008). Furthermore, *pax3b* transcription is dorsally restricted within the neural tube at 24 hpf and does not extend to the anterior limit of *pax3a* expression in the midbrain (Minchin and Hughes, 2008). We sought to investigate the molecular basis of these differing expression profiles by comparing the architecture of CRMs associated with the zebrafish Pax3 paralogs and their *in-vivo* activity. Our comparative genomics study revealed that conserved elements can be readily identified in alignments containing the *pax3a* locus, however inclusion of *pax3b* sequence precluded the discovery of CNEs, discussed in Chapter 5.1. These data indicate that the regulatory sequences associated with the zebrafish Pax3 paralogs have radically diverged following the whole genome duplication in the teleost lineage. Our findings are in contrast to previous studies of the paralogous Pax2 loci in the zebrafish and fugu genomes, which share more than 40 common CNEs (Goode et al., 2011 ; Woofe and Elgar, 2007).

BLAST searches of *pax3b* intron 4 against Condor CNEs revealed a 38 bp fragment that is homologous to Motif1 of Pax3-CNE1 (Fig. 7.11). This region was not detected during our initial comparative genomic alignments as the parameters employed excluded all matches of less than 40 bp. We demonstrate that *pax3b* CNE1 contains regions of measurable homology to all of the regulatory motifs that define this enhancer, however only Motif1 is phylogenetically conserved (Fig. 7.11). *Pax3b*-CNE1 retains its activity within neural progenitors *in-vivo*, demonstrating that despite the altered structure of motifs within this CRM, it is sufficient to recapitulate the primary function of the canonical enhancer (Fig. 7.12). This observation can be attributed to the high degree to which the Pax3/7 PD binding site represented by Motif1 is conserved in the *pax3b* locus (Fig. 7.11). The observed lack of enhancer specificity for neural progenitors and the absence of transgene activation in anterior regions of the CNS may be a result of the poor conservation of the remaining motifs, particularly Motif3 which appears to be required for the activity of *pax3a*-CNE1 *in-vivo*.

Our analysis of the zebrafish Pax3 paralogs reveals that CNE1 is the only conserved enhancer associated with the *pax3b* locus. Furthermore, we demonstrate that only the PD binding site represented by Motif1 is phylogenetically conserved between the Pax3 paralogs and higher vertebrates. These data imply that the initiation of *pax3b* expression in the zebrafish neural plate is dependant on positive feedback from *pax3a*,

which is expressed earlier due to the activity of CNE3. Once initiated, *pax3b* expression is subject to direct autoregulation and positive feedback from *pax3a*, *pax7a*, and *pax7b*. Together, our findings indicate that the lack of CNE3 conservation within the *pax3b* locus and the divergent architecture of CNE1 might be the main contributor to the observed differences in the spatiotemporal expression patterns of the zebrafish Pax3 paralogs (Minchin and Hughes, 2008).

9.7 – The pattern generating properties of autoregulation and positive feedback

Our findings demonstrate that Pax3 expression is subject to direct positive autoregulation and positive feedback from Pax7 during CNS development. This mechanism of regulation is novel in the context of progenitor markers within the neural tube GRN, as all previous genetic interactions have been shown to facilitate transcriptional repression (Dessaud et al., 2008). However, previous reports have highlighted the properties conferred by autoregulatory and positive feedback mechanisms that are relevant to the coordination of transcriptional programmes during embryonic patterning.

One of the best-described roles of developmental autoregulation is the maintenance of transcription in response to transient inductive signals, such as morphogens. This mechanism is exemplified by the regulation of the *Drosophila* gene *Single-minded* (*sim*), which is required for the correct development of the CNS (Nambu et al., 1990). *Sim* expression is induced in a defined stripe within the neurogenic ectoderm by the coordinated activities of the Dorsal gradient and the Notch signaling pathway (Hong et al., 2008). Accordingly, *sim* mutant embryos assessed at early stages of development do not exhibit a loss of *sim* transcription (Nambu et al., 1991). However at later embryonic stages, *sim* mutants reveal a striking loss of RNA transcription, indicating that positive autoregulation is required to transform the transient input of the Dorsal gradient into a stable expression state (Nambu et al., 1991). This conclusion is supported by data demonstrating that *sim* protein directly interacts with its own regulatory element specifically at late stages of *Drosophila* development (Muralidhar et al., 1993). Similar mechanisms, in which autoregulation maintains the transcriptional response to transient cues, have been shown to regulate the expression of terminal selector genes in nematodes (Hobert, 2011), *glial cells missing* during sea urchin development (Ransick and Davidson, 2006 ; Ransick and Davidson, 2012) and *Distalless* during *Drosophila* leg development (Estella et al., 2008).

In addition to establishing a transcriptional memory of cell identity, direct autoregulation has been shown to increase the expression level of key fate determinants during development. This point is illustrated by studies analysing the regulation *fushi tarazu* expression during *Drosophila* segmentation. Reporter gene expression driven by an upstream enhancer recapitulates gene expression during development and is lost in mutant embryos, suggesting that this region mediates positive autoregulation (Pick et al., 1990). The *fushi tarazu* autoregulatory enhancer contains 6 putative homeodomain binding sites, none of which are individually required for its activity *in-vivo* (Schier and Gehring, 1992). However, serial mutation of multiple binding sites results in a stepwise decrease in enhancer activity, suggesting that the primary function of autoregulation in this context is to increase the expression level of *fushi tarazu* during the process of *Drosophila* segmentation (Schier and Gehring, 1992).

Finally, autoregulation has been shown to be capable of producing and maintaining sharp gene expression boundaries when coupled with positive feedback. A well-characterised example of this mechanism is provided by *Hoxb-4* expression during hindbrain patterning. At early stages of development, a diffuse domain of *Hoxb-4* expression is induced by a gradient of RA that emanates from the adjacent mesoderm (Gould et al., 1998). Once initiated, Hoxb-4 direct autoregulation acts to both refine and maintain the initial domain of expression specified by RA (Gould et al., 1997). Moreover, Hoxb-4 protein directly regulates the expression of the RA pathway effector, *RARβ,* in the hindbrain (Serpente et al., 2005). Thus, *Hoxb-4* expression is both dependent on the RA signaling pathway for its activation and positively regulates one of its major transcriptional effectors at later stages of development. This complex regulatory circuit acts to generate and maintain a clearly delineated domain of *Hoxb-4* expression in response to the graded activity of the RA signaling pathway.

Taken together, the results of these studies demonstrate the ability of transcriptional autoregulation to convert transient inductive signals into stable expression states. Furthermore, these data demonstrate that diffuse gene expression boundaries can be refined and maintained throughout development by transcriptional circuits in which autoregulation is coupled to positive feedback. Given our analysis of CNE3 activity during CNS development and the temporal activity of the Wnt signaling pathway during this process, these pattern generating properties of transcriptional autoregulation are likely to be essential for the maintenance and refinement of the Pax3 domain during neural tube development.

9.8 – Conclusions

The results obtained from this study support the established role of the Wnt signaling pathway in the induction of Pax3 expression in the neural plate via CNE3. However, this investigation is the first to describe the role of transcriptional repressors in this process. Furthermore, our data demonstrate that Pax3 transcription is subject to direct autoregulation and positive feedback from Pax7 via CNE1, a novel CNS specific regulatory element. Taken together, our findings indicate that the temporal activity of two distinct enhancers underlies the regulatory logic of Pax3 expression in the neural tube.

In this model, the expression of Pax3 is initiated in the CNS by the direct binding of Wnt signaling pathway effectors to the enhancer region represented by ECR2/CNE3 (Fig. 9.3A). This conclusion is supported by the observations that Tcf3 is bound to this regulatory region in mouse ESC and that mutation of Tcf/Lef sites within this sequence abolishes its activity *in-vivo* (Fig. 8.3) (Degenhardt et al., 2010 ; Marson et al., 2008). However, analysis of transgenic Wnt reporters demonstrates that this pathway alone does not provide sufficient positional information to induce Pax3 expression in the lateral neural plate (Figs. 8.4 and 9.3A) (Dorsky et al., 2002). Our findings demonstrate that two highly conserved motifs within CNE3, named Motif3 and Motif5, are required for the repression of transgene activation in the medial regions of the neural plate (Fig. 8.5). Furthermore, we demonstrate that these motifs contain phylogenetically conserved homeobox and Fox binding sites that are both necessary for the repression of CNE3 activity *in-vivo* and sufficient for Nkx6.1 and Foxa2 binding *in-vitro* (Table 8.1 and Figs. 8.6 and 8.7). Thus, our data suggest that the induction of Pax3 transcription in the lateral neural plate is mediated by the direct binding of Wnt pathway effectors and transcriptional repressors to ECR2/CNE3 (Fig. 9.3A).

Once induced, Pax3 transcription in the neural plate becomes subject to direct autoregulation (Fig. 9.3B). This is evident in our data as CNE1 transcriptional activity in Pax3 expressing cells within the lateral neural plate of zebrafish embryos assessed at bud stage (Fig. 9.3B). Consequently at early stages of CNS patterning, Pax3 expression is regulated by Wnt mediated activation throughout the tissue and direct repression medially via CNE3. Cells in the lateral neural plate that express Pax3 reinforce their transcriptional identity by CNE1 mediated autoregulation (Fig. 9.3B). We propose that this complex regulatory circuit acts to establish the position of the Pax3 expression domain in the neural plate. This hypothesis is supported by our analysis of

Pax3$^{Cre/+}$*/Rosa26-YFP* transgenic mice, which demonstrates that the Pax3 domain is established by a switch in cell identity before E9.5 and subsequently maintained at this position throughout development, discussed in Chapter 3.

The transcriptional activity of the Wnt pathway decreases in the neural tube as development progresses, evident *in-vivo* as a marked downregulation of Wnt reporter transgene expression prior to the maxima of Pax3 transcription in neural tube progenitors (Fig. 8.4) (Dorsky et al., 2002 ; Maretto et al., 2003). Similarly, putative HD and Fox transcriptional repressors do not share a boundary with Pax3 following the closure of the neural tube (Dessaud et al., 2008 ; Jessell, 2000 ; Lewis, 2006). These data indicate that the regulatory interactions mediated by CNE3 are transient and do not function to maintain Pax3 expression at later stages of neural tube patterning. Consequently, we propose that CNE1 mediated autoregulation coupled to Pax7 positive feedback acts to maintain the position of Pax3 ventral boundary in the absence of external transcriptional inputs following neurulation (Fig. 9.3C). This conclusion is supported by the respective profiles of CNE1 and CNE3 activity *in-vivo* and the finding that CNE1 is the only Pax3 enhancer associated with p300 binding in CNS derived tissues at E11.5 (Visel et al., 2009). Our data demonstrates that the Pax3-GFP^{I150} stable line recapitulates the expression profile of *pax3a* in the spinal cord (Fig. 4.7) and furthermore, that CNE1 and CNE3 are the only enhancers contained within this genomic interval that are active in this progenitor domain. These results strongly suggest that the combinatorial activity of these CRMs is sufficient to induce and maintain Pax3 expression in the dorsal spinal cord.

The primary function of developmental patterning is the assignment and maintenance of a unique molecular identity within a defined population of progenitors. This process requires the translation of transient inductive cues, such as morphogens, into bistable expression domains. Our studies suggest that the regulatory logic governing Pax3 expression is based upon the spatial interpretation of transient external inputs and the maintenance of transcription by PD mediated autoregulation and positive feedback. By contrast, the ventral neural tube GRN operates by the progressive assignment of cell identity over time and the refinement of expression boundaries by HD and bHLH mediated repression. Despite their divergent regulatory architectures, we propose that these transcriptional networks are capable of performing similar patterning functions.

Bistability is ensured by the ventral neural tube GRN by a regulatory logic that requires cells to make a binary choice of identity. For example, cells at the p3/pMN

boundary can either transcribe Nkx2.2 or Olig2 but are precluded from expressing both fate determinants (Balaskas et al., 2012). Similarly, the regulatory circuit that coordinates Pax3 transcription following neurulation is capable of defining bistable states by requiring cells to express both Pax3 and Pax7 or neither factor. Autoregulation coupled to positive feedback might also act to refine and maintain the ventral boundary of Pax3 expression in the neural tube in the absence of transcriptional repression, consistent with reports examining the regulation of *Hoxb-4* expression in the hindbrain (Gould et al., 1997 ; Serpente et al., 2005). In the ventral neural tube GRN, fluctuations in the level of Shh signaling are buffered by a series of complex cross-repressive interactions that confers robustness to developmental patterning (Balaskas et al., 2012). We propose that Pax3 autoregulation might provide a mechanism of compensating for changes in the level of Wnt signaling during the early stages of neural plate patterning by enabling cells to maintain their transcriptional identity. Furthermore, our findings suggest that autoregulation coupled to positive feedback might be sufficient to maintain Pax3 expression independently of external inputs at later stages of development. Thus despite its novel architecture in the context of neural tube patterning, the regulatory logic that coordinates Pax3 expression is capable of both defining and maintaining cell identity in response to transient inductive cues.

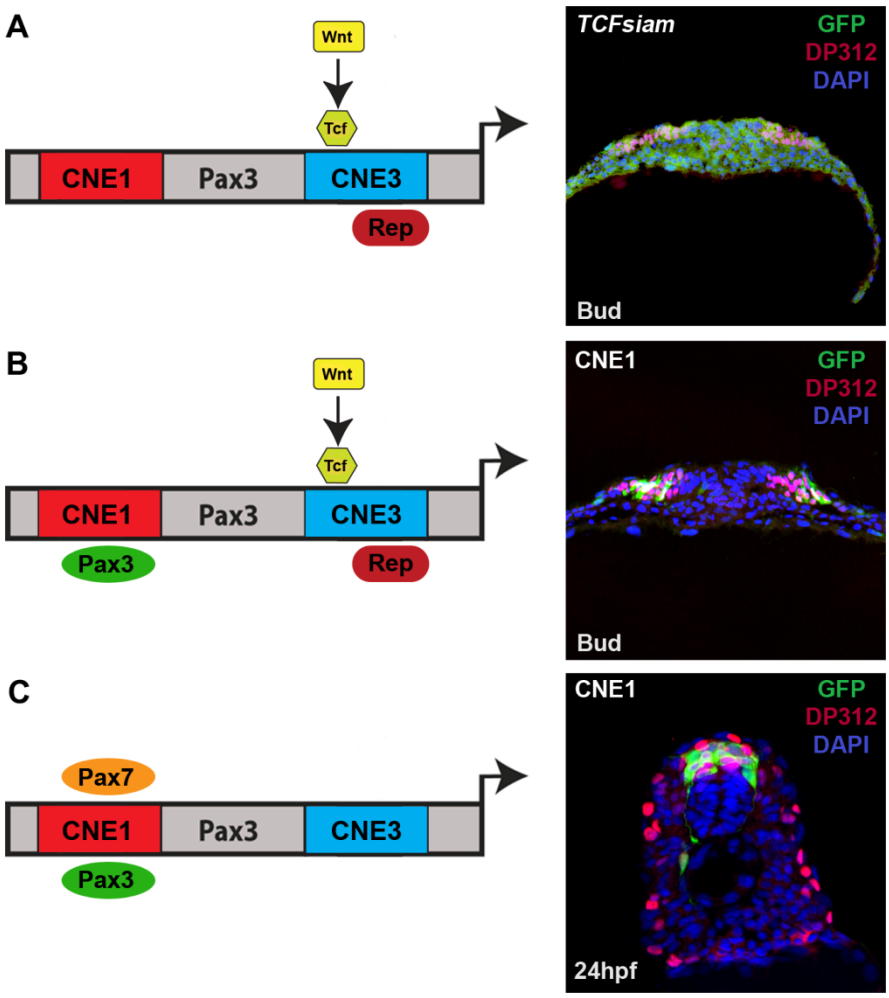


Fig. 9.3 – The regulatory logic of Pax3 expression in the neural tube

The induction of Pax3 transcription in the neural plate is mediated by the direct binding of Wnt pathway effectors, such as Tcf3, to the enhancer region represented by ECR2/CNE3 (A). Our analyses of *TCFsiam* embryos reveals that the transcriptional activity of the Wnt pathway is not patterned across the medio-lateral axis of the neural plate, whereas Pax3 expression is restricted to the intermediate and lateral regions of the tissue (A). Our data suggest that the direct binding of putative HD and Fox family transcriptional repressors to ECR2/CNE3 patterns the Wnt mediated induction of Pax3 expression the neural plate (A). Once expressed, Pax3 transcription is subject to direct autoregulation mediated by a Pax3/7 PD site within CNE1 (B). Consequently, Pax3 expression in the neural plate is regulated by the Wnt pathway, transcriptional repression and autoregulation (B) Following neurulation, the Pax3 expression domain is maintained by CNE1 mediated autoregulation and positive feedback from Pax7 in the absence of CNE3 mediated transcriptional inputs (C).

9.9 – Future work

The findings of this study provide novel insights into the formation of the Pax3 expression domain within the neural tube, the cellular lineage of V0 interneurons and the molecular basis of Pax3 regulation in the CNS. A series of further experiments based on these results would both strengthen the conclusions of this study and provide data connecting Pax3 function to the neural tube GRN.

Our lineage trace data demonstrates that the position of the Pax3 domain within the neural tube is established before E9.5 in mice by a transcriptional switch in cell identity (Fig. 3.2). The findings suggest that transcription is induced within the intermediate region of the neural plate and subsequently becomes restricted to the neural folds, in contrast to the results of previous expression studies (Goulding et al., 1991). In order to validate this hypothesis, future experiments could determine the position of Pax3 expressing cells relative to their lineage at early stages of mouse neural plate patterning. Furthermore, our lineage study reveals that cell migration within the DV axis of the mouse neuroepithelium is temporally restricted with similar dynamics to those observed in chick embryos (Erskine et al., 1998 ; Leber and Sanes, 1995). Pax3/7 have been implicated in this process, due to their ability to directly regulate *Cadherin-7* expression at the sulcus limitans via a conserved enhancer (Luo et al., 2006 ; Prasad and Paulson, 2011). However, the expression profiles of Cadherin family members within the mouse neural tube do not appear to correlate with those observed in chick, zebrafish and *Xenopus* embryos (Espeseth et al., 1995 ; Liu et al., 2007 ; Nakagawa and Takeichi, 1995 ; Nakagawa and Takeichi, 1998 ; Liu et al., 2007). Future studies employing *Pax3Cre/Rosa26-YFP* transgenic mice might seek to resolve these discrepancies by assessing the restriction of cell migration in the DV axis of the neural tube relative to the spatiotemporal expression profiles of Cadherin family members.

Functional dissection of CNE3 indicates that the Wnt mediated initiation of Pax3 transcription is modulated by repressors that are expressed in the medial region of the neural plate, discussed in Chapter 8. Our data demonstrates that both Motif3 and Motif5 are required for the repression of CNE3 mediated transcription from the medial region of the neural plate (Fig. 8.5). Furthermore, our findings reveal that conserved homeobox and Fox binding sites within these Motifs facilitate the binding of putative repressors to CNE3 (Figs. 8.5 and 8.8). We demonstrate that both Nkx6.1 and Foxa2 can bind to DNA spanning these regulatory motifs, however the results of previous investigations do not support the role of the transcription factors in the regulation of

Pax3 expression (Dichmann and Harland, 2010 ; Metzakopian et al., 2012 ; Sander et al., 2000 ; Vallstedt et al., 2001). Further experiments designed to investigate the repression of CNE3 activity from the medial neural plate should be based on the interaction of candidate *in-vitro* synthesised proteins with Motif3 and Motif5 in EMSAs. In addition to Nkx6.1 and Foxa2, these experiments could include Nkx1.2, Cdx-1 and Cdx-2, all of which are homeodomain containing transcription factors that appear to be expressed in the medial neural plate (Bae et al., 2004 ; Marom et al., 1997). The function of putative repressors could be investigated *in-vivo* by ectopic expression of these factors in zebrafish embryos and assaying Pax3 transcription within the neural plate.

Our analysis of a second CNS specific enhancer associated with the Pax3 locus, CNE1, reveals that this CRM requires the presence of two conserved motifs. We demonstrate that Motif1 represents a novel PD binding site that facilitates direct autoregulation and positive feedback, discussed in Chapter 7. Future experiments could further validate these conclusions by ectopically expressing mouse Pax3 or Pax7 in the chick neural tube and assaying the transcription of endogenous Pax3 mRNA. Our findings also reveal that Motif3 is required for CNE1 function *in-vivo* (Fig. 7.2). We were unable to determine the regulatory interactions mediated by CNE1-Motif3 using both *in-vitro* synthesised candidate HMG binding proteins and nuclear extract. In order to advance our analysis of CNE1-Motif3, a series of DNA probes spanning the region should be used in EMSAs with both selected candidate transcription factors and nuclear extract. The addition of these data would complete our functional annotation of CNE1 activity.

Previous reports examining the regulation of Pax3 transcription in the neural tube have highlighted the need to determine the requirement of putative enhancers *in-vivo* (Degenhardt et al., 2010). During the course of this investigation, several attempts were made to assess the requirement of CNE1 and CNE3 for Pax3 expression by deleting these regions from the DKEY-20F20 BAC. We chose to perform these manipulations using the *galK* recombineering protocol, which has previously been successfully employed in the lab (Warming et al., 2005). However, we were unable to confirm the deletion of either CNE1 or CNE3 from DKEY-20F20. Further experiments should aim to assess the requirement of these enhancers for Pax3 expression *in-vivo*, however these studies cannot be performed in mice due to the presence of a redundant CRM located proximally to the promoter (Degenhart et al., 2010 ; Millewski et al., 2004 ; Natoli et al., 1997 ; Pruitt et al., 2004). Consequently, experiments designed to assess

the requirement of CNE1 and CNE3 should be performed in zebrafish embryos and based upon modification of the DKEY-20F20 BAC, achieved either by mutagenic PCR or homology mediated recombination.

Pax3 and Pax7 represent paralogs that are derived from a single ancestral Pax3/7 gene (Balczarek et al., 1997). Consequently, these genes share a high degree of homology in their coding sequences. Despite these similarities, Pax3 and Pax7 exhibit distinct spatiotemporal expression patterns in the neural tube (Goulding et al., 1991 ; Goulding et al., 1993 ; Jostes et al., 1990 ; Kawakami et al., 1997). We sought to investigate the molecular basis of these divergent expression profiles by comparing the activity and regulation of CRMs associated with each locus. Despite employing a similar method of comparative genomics to that used to annotate the Pax3 locus, this study was unable to define any functional enhancers associated with Pax7. Moreover, CNE1 and CNE3 do not appear to be shared by the vertebrate group III Pax genes. Future studies investigating the molecular basis of Pax7 expression could be based on *de-novo* comparative genomic analysis of the locus, substituting *pax7b* sequence for fugu Pax7. These data could then be used in conjunction with p300 binding data and histone methylation marks to define a set of putative Pax7 regulatory elements for further investigation.

Finally, the function of Pax3 and Pax7 in the context of neural tube patterning remains poorly defined. During this investigation, we sought to investigate the function of Pax3 and Pax7 in the neural tube by ectopically expressing either full length, dominant active or dominant negative forms these proteins in chick embryos (Data not shown). The results of these experiments indicated that a high level of ectopic Pax3/7 expression severely affects the cell cycle, neural tube morphology, patterning and differentiation (Data not shown). Consequently, further experiments designed to assess the function of Pax3 and Pax7 in the neural tube should aim to ectopically express transgenes at near endogenous levels. This could be achieved in mice by expressing constructs from either the Pax3 or Pax7 locus and in chick embryos by the use of RCAS retroviral vectors. Together, the results of these investigations will provide data to establish the function of Pax3 and Pax7 in the context of the neural tube GRN.

References

Adams, B., Dorfler, P., Aguzzi, A., Kozmik, Z., Urbanek, P., Maurer-Fogy, I., and Busslinger, M. 1992. Pax-5 encodes the transcription factor BSAP and is expressed in B lymphocytes, the developing CNS, and adult testis. *Genes Dev* **6**: 1589-1607

Ahituv, N., Zhu, Y., Visel, A., Holt, A., Afzal, V., Pennacchio, L.A., and Rubin, E.M. 2007. Deletion of ultraconserved elements yields viable mice. *PLoS Biol* **5**: e234.

Alvarez-Medina, R., Cayuso, J., Okubo, T., Takada, S., and Marti, E. 2008. Wnt canonical pathway restricts graded Shh/Gli patterning activity through the regulation of Gli3 expression. *Development* **135**: 237-247.

Amano, T., Sagai, T., Tanabe, H., Mizushina, Y., Nakazawa, H., and Shiroishi, T. 2009. Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. *Dev Cell* **16**: 47-57.

Aparicio, S., Morrison, A., Gould, A., Gilthorpe, J., Chaudhuri, C., Rigby, P., Krumlauf, R., and Brenner, S. 1995. Detecting conserved regulatory elements with the model genome of the Japanese puffer fish, Fugu rubripes. *Proc Natl Acad Sci USA* **92**: 1684-1688.

Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J.M., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A., Gelpke, M.D., Roach, J., Oh, T., Ho, I.Y., Wong, M., Detter, C., Verhoef, F., Predki, P., Tay, A., Lucas, S., Richardson, P., Smith, S.F., Clark, M.S., Edwards, Y.J., Doggett, N., Zharkikh, A., Tavtigian, S.V., Pruss, D., Barnstead, M., Evans, C., Baden, H., Powell, J., Glusman, G., Rowen, L., Hood, L., Tan, Y.H., Elgar, G., Hawkins, T., Venkatesh, B., Rokhsar, D., and Brenner, S. 2002. Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. *Science* **297**: 1301-1310.

Appel, B. 2000. Zebrafish neural induction and patterning. *Dev Dyn* **219**: 155-168.

Arber, S., Han, B., Mendelsohn, M., Smith, M., Jessell, T.M., and Sockanathan, S. 1999. Requirement for the homeobox gene Hb9 in the consolidation of motor neuron identity. *Neuron* **23**: 659-674.

Auerbach, R. 1954. Analysis of the developmental effects of a lethal mutation in the house mouse. *Journal of Experimental Zoology* **127**:305-329.

Avraham, O., Hadas, Y., Vald, L., Hong, S., Song, M.R., and Klar, A. 2010. Motor and dorsal root ganglion axons serve as choice points for the ipsilateral turning of dI3 axons. *J Neurosci* **30**: 15546-15557.

Bae, Y.K., Shimizu, T., Muraoka, O., Yabe, T., Hirata, T., Nojima, H., Hirano, T., and Hibi, M. 2004. Expression of sax1/nkx1.2 and sax2/nkx1.1 in zebrafish. *Gene Expr Patterns* **4**: 481-486.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**: W202-208.

Balaskas, N., Ribeiro, A., Panovska, J., Dessaud, E., Sasai, N., Page, K.M., Briscoe, J., and Ribes, V. 2012. Gene regulatory logic for reading the Sonic Hedgehog signaling gradient in the vertebrate neural tube. *Cell* **148**: 273-284.

Balczarek, K.A., Lai, Z.C., and Kumar, S. 1997. Evolution of functional diversification of the paired box (Pax) DNA-binding domains. *Mol Biol Evol* **14**: 829-842.

Banerji, J., Rusconi, S., and Schaffner, W. 1981. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**: 299-308.

Banerji, J., Olson, L., and Schaffner, W. 1983. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell* **33**: 729-740.

Bang, A.G., Papalopulu, N., Kintner, C., and Goulding, M.D. 1997. Expression of Pax-3 is initiated in the early neural plate by posteriorizing signals produced by the organizer and by posterior non-axial mesoderm. *Development* **124**: 2075-2085.

Bang, A.G., Papalopulu, N., Goulding, M.D., and Kintner, C. 1999. Expression of Pax-3 in the lateral neural plate is dependent on a Wnt-mediated signal from posterior nonaxial mesoderm. *Dev Biol* **212**: 366-380.

Baumgartner, S., Bopp, D., Burri, M., and Noll, M. 1987. Structure of two genes at the gooseberry locus related to the paired gene and their spatial expression during Drosophila embryogenesis. *Genes Dev* **1**: 1247-1267

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., and Haussler, D. 2004. Ultraconserved elements in the human genome. *Science* **304**: 1321-1325.

Bertrand, N., Castro, D.S., and Guillemot, F. 2002. Proneural genes and the specification of neural cell types. *Nat Rev Neurosci* **3**: 517-530.

Bogdanovic, O., Fernandez-Minan, A., Tena, J.J., de Lacalle-Mustienes, E., Hidalgo, C., van Kruysbergen, I., van Heeringen, S.J., Veenstra, G.J., and Gomez-Skarmeta, J.L. 2012. Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. *Genome Res*.

Bonner, J., Gribble, S.L., Veien, E.S., Nikolaus, O.B., Weidinger, G., and Dorsky, R.I. 2008. Proliferation and patterning are mediated independently in the dorsal spinal cord downstream of canonical Wnt signaling. *Dev Biol* **313**: 398-407.

Bopp, D., Burri, M., Baumgartner, S., Frigerio, G., and Noll, M. 1986. Conservation of a large protein domain in the segmentation gene paired and in functionally related genes of Drosophila. *Cell* **47**: 1033-1040.

Bottcher, R.T. and Niehrs, C. 2005. Fibroblast growth factor signaling during early vertebrate development. *Endocr Rev* **26**: 63-77.

Briscoe, J. and Ericson, J. 2001. Specification of neuronal fates in the ventral neural tube. *Curr Opin Neurobiol* **11**: 43-49.

Briscoe, J., Pierani, A., Jessell, T.M., and Ericson, J. 2000. A homeodomain protein code specifies progenitor cell identity and neuronal fate in the ventral neural tube. *Cell* **101**: 435-445.

Buckingham, M. and Relaix, F. 2007. The role of Pax genes in the development of tissues and organs: Pax3 and Pax7 regulate muscle progenitor cell functions. *Annu Rev Cell Dev Biol* **23**: 645-673

Chalepakis, G., Jones, F.S., Edelman, G.M., and Gruss, P. 1994. Pax-3 contains domains for transcription activation and transcription inhibition. *Proc Natl Acad Sci USA* **91**: 12745-12749.

Chalepakis, G. and Gruss, P. 1995. Identification of DNA recognition sequences for the Pax3 paired domain. *Gene* **162**: 267-270.

Chi, N. and Epstein, J.A. 2002. Getting your Pax straight: Pax proteins in development and disease. *Trends Genet* **18**: 41-47.

Chiang, C., Litingtung, Y., Lee, E., Young, K.E., Corden, J.L., Westphal, H., and Beachy, P.A. 1996. Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. *Nature* **383**: 407-413.

Chamberlain, C.E., Jeong, J., Guo, C., Allen, B.L., and McMahon, A.P. 2008. Notochord-derived Shh concentrates in close association with the apically positioned basal body in neural target cells and forms a dynamic gradient during neural patterning. *Development* **135**: 1097-1106.

Cheesman, S.E., Layden, M.J., Von Ohlen, T., Doe, C.Q., and Eisen, J.S. 2004. Zebrafish and fly Nkx6 proteins have similar CNS expression patterns and regulate motoneuron formation. *Development* **131**: 5221-5232.

Chen, J.A., Huang, Y.P., Mazzoni, E.O., Tan, G.C., Zavadil, J., and Wichterle, H. 2011. Mir-17-3p controls spinal neural progenitor patterning by regulating Olig2/Irx3 cross-repressive loop. *Neuron* **69**: 721-735.

Chesnutt, C., Burrus, L.W., Brown, A.M., and Niswander, L. 2004. Coordinate regulation of neural tube patterning and proliferation by TGFbeta and WNT activity. *Dev Biol* **274**: 334-347.

Copp, A.J., Greene, N.D., and Murdoch, J.N. 2003. The genetic basis of mammalian neurulation. *Nat Rev Genet* **4**: 784-793.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., Boyer, L.A., Young, R.A., and Jaenisch, R. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci USA* **107**: 21931-21936.

Curran, K., Raible, D.W., and Lister, J.A. 2009. Foxd3 controls melanophore specification in the zebrafish neural crest by regulation of Mitf. *Dev Biol* **332**: 408-41

Cvekl, A., Kashanchi, F., Brady, J.N., and Piatigorsky, J. 1999. Pax-6 interactions with TATA-box-binding protein and retinoblastoma protein. *Invest Ophthalmol Vis Sci* **40**: 1343-1350.

Czerny, T., Schaffner, G., and Busslinger, M. 1993. DNA sequence recognition by Pax proteins: bipartite structure of the paired domain and its binding site. *Genes Dev* **7**: 2048-2061.

Davis, G.K., D'Alessio, J.A., and Patel, N.H. 2005. Pax3/7 genes reveal conservation and divergence in the arthropod segmentation hierarchy. *Dev Biol* **285**: 169-184.

Degenhardt, K.R., Milewski, R.C., Padmanabhan, A., Miller, M., Singh, M.K., Lang, D., Engleka, K.A., Wu, M., Li, J., Zhou, D., Antonucci, N., Li, L., and Epstein, J.A. 2010. Distinct enhancers at the Pax3 locus can function redundantly to regulate neural tube and neural crest expressions. *Dev Biol* **339**: 519-527.

de Laat, W., Klous, P., Kooren, J., Noordermeer, D., Palstra, R.J., Simonis, M., Splinter, E., and Grosveld, F. 2008. Three-dimensional organization of gene expression in erythroid cells. *Curr Top Dev Biol* **82**: 117-139.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. 2002. Capturing chromosome conformation. *Science* **295**: 1306-1311.

Dessaud, E., Yang, L.L., Hill, K., Cox, B., Ulloa, F., Ribeiro, A., Mynett, A., Novitch, B.G., and Briscoe, J. 2007. Interpretation of the sonic hedgehog morphogen gradient by a temporal adaptation mechanism. *Nature* **450**: 717-720.

Dessaud, E., McMahon, A.P., and Briscoe, J. 2008. Pattern formation in the vertebrate neural tube: a sonic hedgehog morphogen-regulated transcriptional network. *Development* **135**: 2489-2503.

Dessaud, E., Ribes, V., Balaskas, N., Yang, L.L., Pierani, A., Kicheva, A., Novitch, B.G., Briscoe, J., and Sasai, N. 2010. Dynamic assignment and maintenance of positional identity in the ventral neural tube by the morphogen sonic hedgehog. *PLoS Biol* **8**: e1000382.

Deutsch, U., Dressler, G.R., and Gruss, P. 1988. Pax 1, a member of a paired box homologous murine gene family, is expressed in segmented structures during development. *Cell* **53**: 617-625.

Dichmann, D.S. and Harland, R.M. 2010. Nkx6 genes pattern the frog neural plate and Nkx6.1 is necessary for motoneuron axon projection. *Dev Biol* **349**: 378-386.

Dorsky, R.I., Sheldahl, L.C., and Moon, R.T. 2002. A transgenic Lef1/beta-catenin-dependent reporter is expressed in spatially restricted domains throughout zebrafish development. *Dev Biol* **241**: 229-237.

Dude, C.M., Kuan, C.Y., Bradshaw, J.R., Greene, N.D., Relaix, F., Stark, M.R., and Baker, C.V. 2009. Activation of Pax3 target genes is necessary but not sufficient for neurogenesis in the ophthalmic trigeminal placode. *Dev Biol* **326**: 314-326.

Durston, A.J., Timmermans, J.P., Hage, W.J., Hendriks, H.F., de Vries, N.J., Heideveld, M., and Nieuwkoop, P.D. 1989. Retinoic acid causes an anteroposterior transformation in the developing central nervous system. *Nature* **340**: 140-144.

Eberhard, D. and Busslinger, M. 1999. The partial homeodomain of the transcription factor Pax-5 (BSAP) is an interaction motif for the retinoblastoma and TATA-binding proteins. *Cancer Res* **59**: 1716s-1724s; discussion 1724s-1725s.

Eberhard, D., Jimenez, G., Heavey, B., and Busslinger, M. 2000. Transcriptional repression by Pax5 (BSAP) through interaction with corepressors of the Groucho family. *EMBO J* **19**: 2292-2303.

Elliott, D.A. and Brand, A.H. 2008. The GAL4 system : a versatile system for the expression of genes. *Methods Mol Biol* **420**: 79-95.

Engleka, K.A., Gitler, A.D., Zhang, M., Zhou, D.D., High, F.A., and Epstein, J.A. 2005. Insertion of Cre into the Pax3 locus creates a new allele of Splotch and identifies unexpected Pax3 derivatives. *Dev Biol* **280**: 396-406.

Engstrom, P.G., Fredman, D., and Lenhard, B. 2008. Ancora: a web resource for exploring highly conserved noncoding elements and their association with developmental regulatory genes. *Genome Biol* **9**: R34.

Epstein, D.J., Vekemans, M., and Gros, P. 1991. Splotch (Sp2H), a mutation affecting development of the mouse neural tube, shows a deletion within the paired homeodomain of Pax-3. *Cell* **67**: 767-774.

Epstein, D.J., Vogan, K.J., Trasler, D.G., and Gros, P. 1993. A mutation within intron 3 of the Pax-3 gene produces aberrantly spliced mRNA transcripts in the splotch (Sp) mouse mutant. *Proc Natl Acad Sci USA* **90**: 532-536.

Epstein, D.J., McMahon, A.P., and Joyner, A.L. 1999. Regionalization of Sonic hedgehog transcription along the anteroposterior axis of the mouse central nervous system is regulated by Hnf3-dependent and -independent mechanisms. *Development* **126**: 281-292.

Epstein, J., Cai, J., Glaser, T., Jepeal, L., and Maas, R. 1994. Identification of a Pax paired domain recognition sequence and evidence for DNA-dependent conformational changes. *J Biol Chem* **269**: 8355-8361.

Epstein, J.A., Shapiro, D.N., Cheng, J., Lam, P.Y., and Maas, R.L. 1996. Pax3 modulates expression of the c-Met receptor during limb muscle development. *Proc Natl Acad Sci USA* **93**: 4213-4218.

Ericson, J., Morton, S., Kawakami, A., Roelink, H., and Jessell, T.M. 1996. Two critical periods of Sonic Hedgehog signaling required for the specification of motor neuron identity. *Cell* **87**: 661-673.

Ericson, J., Rashbass, P., Schedl, A., Brenner-Morton, S., Kawakami, A., van Heyningen, V., Jessell, T.M., and Briscoe, J. 1997. Pax6 controls progenitor cell identity and neuronal fate in response to graded Shh signaling. *Cell* **90**: 169-180.

Erskine, L., Patel, K., and Clarke, J.D. 1998. Progenitor dispersal and the origin of early neuronal phenotypes in the chick embryo spinal cord. *Dev Biol* **199**: 26-41.

Espeseth, A., Johnson, E., and Kintner, C. 1995. Xenopus F-cadherin, a novel member of the cadherin family of cell adhesion molecules, is expressed at boundaries in the neural tube. *Mol Cell Neurosci* **6**: 199-211.

Espeseth, A., Marnellos, G., and Kintner, C. 1998. The role of F-cadherin in localizing cells during neural tube formation in Xenopus embryos. *Development* **125**: 301-312.

Estella, C., McKay, D.J., and Mann, R.S. 2008. Molecular integration of wingless, decapentaplegic, and autoregulatory inputs into Distalless during Drosophila leg development. *Dev Cell* **14**: 86-96.

Feng, X., Adiarte, E.G., and Devoto, S.H. 2006. Hedgehog acts directly on the zebrafish dermomyotome to promote myogenic differentiation. *Dev Biol* **300**: 736-746.

Friedrich, G. and Soriano, P. 1991. Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev* **5**: 1513-1523.

Frigerio, G., Burri, M., Bopp, D., Baumgartner, S., and Noll, M. 1986. Structure of the segmentation gene paired and the Drosophila PRD gene set as part of a gene network. *Cell* **47**: 735-746.

Frost, V., Grocott, T., Eccles, M.R., and Chantry, A. 2008. Self-regulated Pax gene expression and modulation by the TGFbeta superfamily. *Crit Rev Biochem Mol Biol* **43**: 371-391.

Garnett, A.T., Square, T.A., and Medeiros, D.M. 2012. BMP, Wnt and FGF signals are integrated through evolutionarily conserved enhancers to achieve robust expression of Pax3 and Zic genes at the zebrafish neural plate border. *Development* **139**: 4220-4231.

Gillies, S.D., Morrison, S.L., Oi, V.T., and Tonegawa, S. 1983. A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell* **33**: 717-728.

Glaser, T., Jepeal, L., Edwards, J.G., Young, S.R., Favor, J., and Maas, R.L. 1994. PAX6 gene dosage effect in a family with congenital cataracts, aniridia, anophthalmia and central nervous system defects. *Nat Genet* **7**: 463-471.

Goetz, S.C. and Anderson, K.V. 2010. The primary cilium: a signaling centre during vertebrate development. *Nat Rev Genet* **11**: 331-344.

Gomez-Skarmeta, J.L., Lenhard, B., and Becker, T.S. 2006. New technologies, new findings, and new concepts in the study of vertebrate cis-regulatory sequences. *Dev Dyn* **235**: 870-885.

Goode, D.K., Callaway, H.A., Cerda, G.A., Lewis, K.E., and Elgar, G. 2011. Minor change, major difference: divergent functions of highly conserved cis-regulatory elements subsequent to whole genome duplication events. *Development* **138**: 879-884.

Gould, A., Morrison, A., Sproat, G., White, R.A., and Krumlauf, R. 1997. Positive cross-regulation and enhancer sharing: two mechanisms for specifying overlapping Hox expression patterns. *Genes Dev* **11**: 900-913.

Gould, A., Itasaki, N., and Krumlauf, R. 1998. Initiation of rhombomeric Hoxb4 expression requires induction by somites and a retinoid pathway. *Neuron* **21**: 39-51.

Goulding, M.D., Chalepakis, G., Deutsch, U., Erselius, J.R., and Gruss, P. 1991. Pax-3, a novel murine DNA binding protein expressed during early neurogenesis. *EMBO J* **10**: 1135-1147.

Goulding, M.D., Lumsden, A., and Gruss, P. 1993. Signals from the notochord and floor plate regulate the region-specific expression of two Pax genes in the developing spinal cord. *Development* **117**: 1001-1016.

Goulding, M., Lanuza, G., Sapir, T., and Narayan, S. 2002. The formation of sensorimotor circuits. *Curr Opin Neurobiol* **12**: 508-515.

Gomez-Skarmeta, J.L., Lenhard, B., and Becker, T.S. 2006. New technologies, new findings, and new concepts in the study of vertebrate cis-regulatory sequences. *Dev Dyn* **235**: 870-885.

Grabher, C., Joly, J.S., and Wittbrodt, J. 2004. Highly efficient zebrafish transgenesis mediated by the meganuclease I-SceI. *Methods Cell Biol* **77**: 381-401.

Gross, M.K., Dottori, M., and Goulding, M. 2002. Lbx1 specifies somatosensory association interneurons in the dorsal spinal cord. *Neuron* **34**: 535-549.

Guner, B. and Karlstrom, R.O. 2007. Cloning of zebrafish nkx6.2 and a comprehensive analysis of the conserved transcriptional response to Hedgehog/Gli signaling in the zebrafish neural tube. *Gene Expr Patterns* **7**: 596-605.

Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S. 2007. Quantifying similarity between motifs. *Genome Biol* **8**: R24.

Hamburger, V. and Hamilton, H.L. 1992. A series of normal stages in the development of the chick embryo. 1951. *Dev Dyn* **195**: 231-272.

Hammond, C.L., Hinits, Y., Osborn, D.P., Minchin, J.E., Tettamanti, G., and Hughes, S.M. 2007. Signals and myogenic regulatory factors restrict pax3 and pax7 expression to dermomyotome-like tissue in zebrafish. *Dev Biol* **302**: 504-521.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., Wang, W., Weng, Z., Green, R.D., Crawford, G.E., and Ren, B. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**: 311-318.

Helms, A.W. and Johnson, J.E. 1998. Progenitors of dorsal commissural interneurons are defined by MATH1 expression. *Development* **125**: 919-928.

Henke, R.M., Meredith, D.M., Borromeo, M.D., Savage, T.K., and Johnson, J.E. 2009. Ascl1 and Neurog2 form novel complexes and regulate Delta-like3 (Dll3) expression in the neural tube. *Dev Biol* **328**: 529-540.

His, W. 1888. Zur Geschichte des Gehirns; sowie der centralen peripherischen Nervenbahnen beim menschlichen Embryo. *Abh. K. Sachs. Ges. Wiss. Math.-Phys. Kl.* **14**: 341-372.

Hobert, O. 2011. Regulation of terminal differentiation programs in the nervous system. *Annu Rev Cell Dev Biol* **27**: 681-696.

Hollyday, M., McMahon, J.A., and McMahon, A.P. 1995. Wnt expression patterns in chick embryo nervous system. *Mech Dev* **52**: 9-25.

Hong, J.W., Hendrix, D.A., Papatsenko, D., and Levine, M.S. 2008. How the Dorsal gradient works: insights from postgenome technologies. *Proc Natl Acad Sci USA* **105**: 20072-20076.

Hoover, A.N., Wynkoop, A., Zeng, H., Jia, J., Niswander, L.A., and Liu, A. 2008. C2cd3 is required for cilia formation and Hedgehog signaling in mouse. *Development* **135**: 4049-4058.

Huangfu, D., Liu, A., Rakeman, A.S., Murcia, N.S., Niswander, L., and Anderson, K.V. 2003. Hedgehog signaling in the mouse requires intraflagellar transport proteins. *Nature* **426**: 83-87.

Hynes, M., Ye, W., Wang, K., Stone, D., Murone, M., Sauvage, F., and Rosenthal, A. 2000. The seven-transmembrane receptor smoothened cell-autonomously induces multiple ventral cell types. *Nat Neurosci* **3**: 41-46.

Ichi, S., Boshnjaku, V., Shen, Y.W., Mania-Farnell, B., Ahlgren, S., Sapru, S., Mansukhani, N., McLone, D.G., Tomita, T., and Mayanil, C.S. 2010. Role of Pax3 acetylation in the regulation of Hes1 and Neurog2. *Mol Biol Cell* **22**: 503-512.

Ingham, P.W. and McMahon, A.P. 2001. Hedgehog signaling in animal development: paradigms and principles. *Genes Dev* **15**: 3059-3087.

Jessell, T.M. 2000. Neuronal specification in the spinal cord: inductive signals and transcriptional codes. *Nat Rev Genet* **1**: 20-29.

Jiang, J., Kosman, D., Ip, Y.T., and Levine, M. 1991. The dorsal morphogen gradient regulates the mesoderm determinant twist in early Drosophila embryos. *Genes Dev* **5**: 1881-1891.

Jostes, B., Walther, C., and Gruss, P. 1990. The murine paired box gene, Pax7, is expressed specifically during the development of the nervous and muscular system. *Mech Dev* **33**: 27-37.

Ju, M.J., Aroca, P., Luo, J., Puelles, L., and Redies, C. 2004. Molecular profiling indicates avian branchiomotor nuclei invade the hindbrain alar plate. *Neuroscience* **128**: 785-796.

Jun, S. and Desplan, C. 1996. Cooperative interactions between paired domain and homeodomain. *Development* **122**: 2639-2650.

Kageyama, R. and Ohtsuka, T. 1999. The Notch-Hes pathway in mammalian neural development. *Cell Res* **9**: 179-188.

Kaneko, K.J., Kohn, M.J., Liu, C., and DePamphilis, M.L. 2007. Transcription factor TEAD2 is involved in neural tube closure. *Genesis* **45**: 577-587.

Kawakami, A., Kimura-Kawakami, M., Nomura, T., and Fujisawa, H. 1997. Distributions of PAX6 and PAX7 proteins suggest their involvement in both early and late phases of chick brain development. *Mech Dev* **66**: 119-130.

Kawakami, K. 2007. Tol2: a versatile gene transfer vector in vertebrates. *Genome Biol* **8 Suppl 1**: S7.

Kiecker, C. and Niehrs, C. 2001. A morphogen gradient of Wnt/beta-catenin signaling regulates anteroposterior neural patterning in Xenopus. *Development* **128**: 4189-4201.

Kilchherr, F., Baumgartner,S., Bopp, D., Frei, E., and Noll, M. 1986 Isolation of the *paired* gene of *Drosophila* and its spatial expression during early embryogenesis. *Nature* **321**: 493-499.

Kimmel, C.B., Warga, R.M., and Kane, D.A. 1994. Cell cycles and clonal strings during formation of the zebrafish central nervous system. *Development* **120**: 265-276.

Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B., and Schilling, T.F. 1995. Stages of embryonic development of the zebrafish. *Dev Dyn* **203**: 253-310.

Kleinjan, D.A. and van Heyningen, V. 2005. Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am J Hum Genet* **76**: 8-32.

Koblar, S.A., Murphy, M., Barrett, G.L., Underhill, A., Gros, P., and Bartlett, P.F. 1999. Pax-3 regulates neurogenesis in neural crest-derived precursor cells. *J Neurosci Res* **56**: 518-530.

Koudijs, M.J., den Broeder, M.J., Groot, E., and van Eeden, F.J. 2008. Genetic analysis of the two zebrafish patched homologues identifies novel roles for the hedgehog signaling pathway. *BMC Dev Biol* **8**: 15.

Ladle, D.R., Pecho-Vrieseling, E., and Arber, S. 2007. Assembly of motor circuits in the spinal cord: driven to function by genetic and experience-dependent mechanisms. *Neuron* **56**: 270-283.

Lang, D., Lu, M.M., Huang, L., Engleka, K.A., Zhang, M., Chu, E.Y., Lipner, S., Skoultchi, A., Millar, S.E., and Epstein, J.A. 2005. Pax3 functions at a nodal point in melanocyte stem cell differentiation. *Nature* **433**: 884-887.

Lanuza, G.M., Gosgnach, S., Pierani, A., Jessell, T.M., and Goulding, M. 2004. Genetic identification of spinal interneurons that coordinate left-right locomotor activity necessary for walking movements. *Neuron* **42**: 375-386.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., and Higgins, D.G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947-2948.

Le Dreau, G., Garcia-Campmany, L., Rabadan, M.A., Ferronha, T., Tozer, S., Briscoe, J., and Marti, E. 2012. Canonical BMP7 activity is required for the generation of discrete neuronal populations in the dorsal spinal cord. *Development* **139**: 259-268.

Leber, S.M., Breedlove, S.M., and Sanes, J.R. 1990. Lineage, arrangement, and death of clonally related motoneurons in chick spinal cord. *J Neurosci* **10**: 2451-2462.

Leber, S.M. and Sanes, J.R. 1995. Migratory paths of neurons and glia in the embryonic chick spinal cord. *J Neurosci* **15**: 1236-1248.

Lechner, M.S. and Dressler, G.R. 1996. Mapping of Pax-2 transcription activation domains. *J Biol Chem* **271**: 21088-21093.

Lee, H.K. and Deneen, B. Daam2 is required for dorsal patterning via modulation of canonical Wnt signaling in the developing spinal cord. *Dev Cell* **22**: 183-196.

Lee, K.J., Dietrich, P., and Jessell, T.M. 2000. Genetic ablation reveals that the roof plate is essential for dorsal interneuron specification. *Nature* **403**: 734-740.

Lettice, L.A., Horikoshi, T., Heaney, S.J., van Baren, M.J., van der Linde, H.C., Breedveld, G.J., Joosse, M., Akarsu, N., Oostra, B.A., Endo, N., Shibata, M., Suzuki, M., Takahashi, E., Shinka, T., Nakahori, Y., Ayusawa, D., Nakabayashi, K., Scherer, S.W., Heutink, P., Hill, R.E., and Noji, S. 2002. Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. *Proc Natl Acad Sci USA* **99**: 7548-7553.

Lettice, L.A., Heaney, S.J., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Goode, D., Elgar, G., Hill, R.E., and de Graaff, E. 2003. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* **12**: 1725-1735.

Lewis, K.E. and Eisen, J.S. 2003. From cells to circuits: development of the zebrafish spinal cord. *Prog Neurobiol* **69**: 419-449.

Lewis, K.E. 2006. How do genes regulate simple behaviours? Understanding how different neurons in the vertebrate spinal cord are genetically specified. *Philos Trans R Soc Lond B Biol Sci* **361**: 45-66.

Liem, K.F., Jr., Tremml, G., Roelink, H., and Jessell, T.M. 1995. Dorsal differentiation of neural plate cells induced by BMP-mediated signals from epidermal ectoderm. *Cell* **82**: 969-979.

Liem, K.F., Jr., Tremml, G., and Jessell, T.M. 1997. A role for the roof plate and its resident TGFbeta-related proteins in neuronal patterning in the dorsal spinal cord. *Cell* **91**: 127-138.

Litingtung, Y. and Chiang, C. 2000. Specification of ventral neuron types is mediated by an antagonistic interaction between Shh and Gli3. *Nat Neurosci* **3**: 979-985.

Liu, B., Joel Duff, R., Londraville, R.L., Marrs, J.A., and Liu, Q. 2007. Cloning and expression analysis of cadherin7 in the central nervous system of the embryonic zebrafish. *Gene Expr Patterns* **7**: 15-22.

Logan, C.Y. and Nusse, R. 2004. The Wnt signaling pathway in development and disease. *Annu Rev Cell Dev Biol* **20**: 781-810.

Loots, G.G. and Ovcharenko, I. 2007. Mulan: multiple-sequence alignment to predict functional elements in genomic sequences. *Methods Mol Biol* **395**: 237-254.

Luo, J., Ju, M.J., and Redies, C. 2006. Regionalized cadherin-7 expression by radial glia is regulated by Shh and Pax7 during chicken spinal cord development. *Neuroscience* **142**: 1133-1143.

Maden, M. 2007. Retinoic acid in the development, regeneration and maintenance of the nervous system. *Nat Rev Neurosci* **8**: 755-765.

Mansouri, A., Stoykova, A., Torres, M., and Gruss, P. 1996. Dysgenesis of cephalic neural crest derivatives in Pax7-/- mutant mice. *Development* **122**: 831-838.

Mansouri, A. and Gruss, P. 1998. Pax3 and Pax7 are expressed in commissural neurons and restrict ventral neuronal identity in the spinal cord. *Mech Dev* **78**: 171-178.

Maretto, S., Cordenonsi, M., Dupont, S., Braghetta, P., Broccoli, V., Hassan, A.B., Volpin, D., Bressan, G.M., and Piccolo, S. 2003. Mapping Wnt/beta-catenin signaling during mouse development and in colorectal tumors. *Proc Natl Acad Sci USA* **100**: 3299-3304.

Markstein, M., Markstein, P., Markstein, V., and Levine, M.S. 2002. Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the Drosophila embryo. *Proc Natl Acad Sci USA* **99**: 763-768.

Markstein, M., Zinzen, R., Markstein, P., Yee, K.P., Erives, A., Stathopoulos, A., and Levine, M. 2004. A regulatory code for neurogenic gene expression in the Drosophila embryo. *Development* **131**: 2387-2394.

Marom, K., Shapira, E., and Fainsod, A. 1997. The chicken caudal genes establish an anterior-posterior gradient by partially overlapping temporal and spatial patterns of expression. *Mech Dev* **64**: 41-52.

Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J., Calabrese, J.M., Dennis, L.M., Volkert, T.L., Gupta, S., Love, J., Hannett, N., Sharp, P.A., Bartel, D.P., Jaenisch, R., and Young, R.A. 2008. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134**: 521-533.

Marti, E., Takada, R., Bumcrot, D.A., Sasaki, H., and McMahon, A.P. 1995. Distribution of Sonic hedgehog peptides in the developing chick and mouse embryo. *Development* **121**: 2537-2547.

Massague, J., Seoane, J., and Wotton, D. 2005. Smad transcription factors. *Genes Dev* **19**: 2783-2810.

Metzakopian, E., Lin, W., Salmon-Divon, M., Dvinge, H., Andersson, E., Ericson, J., Perlmann, T., Whitsett, J.A., Bertone, P., and Ang, S.L. 2012. Genome-wide characterization of Foxa2 targets reveals upregulation of floor plate genes and repression of ventrolateral genes in midbrain dopaminergic progenitors. *Development* **139**: 2625-2634.

Meyer, N.P. and Roelink, H. 2003. The amino-terminal region of Gli3 antagonizes the Shh response and acts in dorsoventral fate specification in the developing spinal cord. *Dev Biol* **257**: 343-355.

Miele, A. and Dekker, J. 2008. Long-range chromosomal interactions and gene regulation. *Mol Biosyst* **4**: 1046-1057.

Milewski, R.C., Chi, N.C., Li, J., Brown, C., Lu, M.M., and Epstein, J.A. 2004. Identification of minimal enhancer elements sufficient for Pax3 expression in neural crest and implication of Tead2 as a regulator of Pax3. *Development* **131**: 829-837.

Minchin, J.E. and Hughes, S.M. 2008. Sequential actions of Pax3 and Pax7 drive xanthophore development in zebrafish neural crest. *Dev Biol* **317**: 508-522.

Miura, H., Yanazawa, M., Kato, K., and Kitamura, K. 1997. Expression of a novel aristaless related homeobox gene 'Arx' in the vertebrate telencephalon, diencephalon and floor plate. *Mech Dev* **65**: 99-109.

Moore, R., Champeval, D., Denat, L., Tan, S.S., Faure, F., Julien-Grille, S., and Larue, L. 2004. Involvement of cadherins 7 and 20 in mouse embryogenesis and melanocyte transformation. *Oncogene* **23**: 6726-6735.

Moran-Rivard, L., Kagawa, T., Saueressig, H., Gross, M.K., Burrill, J., and Goulding, M. 2001. Evx1 is a postmitotic determinant of v0 interneuron identity in the spinal cord. *Neuron* **29**: 385-399.

Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M.P., and Chambon, P. 1981. The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Res* **9**: 6047-6068.

Moro, E., Ozhan-Kizil, G., Mongera, A., Beis, D., Wierzbicki, C., Young, R.M., Bournele, D., Domenichini, A., Valdivia, L.E., Lum, L., Chen, C., Amatruda, J.F., Tiso, N., Weidinger, G., and Argenton, F. 2012. In vivo Wnt signaling tracing through a transgenic biosensor fish reveals novel activity domains. *Dev Biol* **366**: 327-340.

Moussian, B. and Roth, S. 2005. Dorsoventral axis formation in the Drosophila embryo--shaping and transducing a morphogen gradient. *Curr Biol* **15**: R887-899.

Muhr, J., Andersson, E., Persson, M., Jessell, T.M., and Ericson, J. 2001. Groucho-mediated transcriptional repression establishes progenitor cell pattern and neuronal fate in the ventral neural tube. *Cell* **104**: 861-873.

Muller, F., Williams, D.W., Kobolak, J., Gauvry, L., Goldspink, G., Orban, L., and Maclean, N. 1997. Activator effect of coinjected enhancers on the muscle-specific expression of promoters in zebrafish embryos. *Mol Reprod Dev* **47**: 404-412.

Muller, F., Chang, B., Albert, S., Fischer, N., Tora, L., and Strahle, U. 1999. Intronic enhancers control expression of zebrafish sonic hedgehog in floor plate and notochord. *Development* **126**: 2103-2116.

Muller, F., Blader, P., and Strahle, U. 2002. Search for enhancers: teleost models in comparative genomic and transgenic analysis of cis regulatory elements. *Bioessays* **24**: 564-572.

Muralidhar, M.G., Callahan, C.A., and Thomas, J.B. 1993. Single-minded regulation of genes in the embryonic midline of the Drosophila central nervous system. *Mech Dev* **41**: 129-138.

Muroyama, Y., Fujihara, M., Ikeya, M., Kondoh, H., and Takada, S. 2002. Wnt signaling plays an essential role in neuronal specification of the dorsal spinal cord. *Genes Dev* **16**: 548-553.

Nakagawa, S. and Takeichi, M. 1995. Neural crest cell-cell adhesion controlled by sequential and subpopulation-specific expression of novel cadherins. *Development* **121**: 1321-1332.

Nakagawa, S. and Takeichi, M. 1998. Neural crest emigration from the neural tube depends on regulated cadherin expression. *Development* **125**: 2963-2971.

Nakazaki, H., Reddy, A.C., Mania-Farnell, B.L., Shen, Y.W., Ichi, S., McCabe, C., George, D., McLone, D.G., Tomita, T., and Mayanil, C.S. 2008. Key basic helix-loop-helix transcription factor genes Hes1 and Ngn2 are regulated by Pax3 during mouse embryonic development. *Dev Biol* **316**: 510-523.

Nambu, J.R., Franks, R.G., Hu, S., and Crews, S.T. 1990. The single-minded gene of Drosophila is required for the expression of genes important for the development of CNS midline cells. *Cell* **63**: 63-75.

Nambu, J.R., Lewis, J.O., Wharton, K.A., Jr., and Crews, S.T. 1991. The Drosophila single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development. *Cell* **67**: 1157-1167.

Natoli, T.A., Ellsworth, M.K., Wu, C., Gross, K.W., and Pruitt, S.C. 1997. Positive and negative DNA sequence elements are required to establish the pattern of Pax3 expression. *Development* **124**: 617-626.

Newburger, D.E. and Bulyk, M.L. 2009. UniPROBE: an online database of protein binding microarray data on protein-DNA interactions. *Nucleic Acids Res* **37**: D77-82.

Niehrs, C. 2004. Regionally specific induction by the Spemann-Mangold organizer. *Nat Rev Genet* **5**: 425-434.

Noonan, J.P. and McCallion, A.S. Genomics of long-range regulatory elements. *Annu Rev Genomics Hum Genet* **11**: 1-23.

Nordstrom, U., Jessell, T.M., and Edlund, T. 2002. Progressive induction of caudal neural character by graded Wnt signaling. *Nat Neurosci* **5**: 525-532.

Novitch, B.G., Chen, A.I., and Jessell, T.M. 2001. Coordinate regulation of motor neuron subtype identity and pan-neuronal properties by the bHLH repressor Olig2. *Neuron* **31**: 773-789.

Nusslein-Volhard, C. and Wieschaus, E. 1980. Mutations affecting segment number and polarity in Drosophila. *Nature* **287**: 795-801.

Ovcharenko, I., Nobrega, M.A., Loots, G.G., and Stubbs, L. 2004. ECR Browser: a tool for visualizing and accessing data from comparisons of multiple vertebrate genomes. *Nucleic Acids Res* **32**: W280-286.

Ovcharenko, I., Loots, G.G., Giardine, B.M., Hou, M., Ma, J., Hardison, R.C., Stubbs, L., and Miller, W. 2005. Mulan: multiple-sequence local alignment and visualization for studying function and evolution. *Genome Res* **15**: 184-194.

Pani, L., Horal, M., and Loeken, M.R. 2002. Rescue of neural tube defects in Pax-3-deficient embryos by p53 loss of function: implications for Pax-3- dependent development and tumorigenesis. *Genes Dev* **16**: 676-680.

Papan, C. and Campos-Ortega, J.A. 1999. Region-specific cell clones in the developing spinal cord of the zebrafish. *Dev Genes Evol* **209**: 135-144.

Papatsenko, D. and Levine, M. 2005. Quantitative analysis of binding motifs mediating diverse spatial readouts of the Dorsal gradient in the Drosophila embryo. *Proc Natl Acad Sci USA* **102**: 4966-4971.

Parr, B.A., Shea, M.J., Vassileva, G., and McMahon, A.P. 1993. Mouse Wnt genes exhibit discrete domains of expression in the early embryonic CNS and limb buds. *Development* **119**: 247-261.

Patel, S.R., Bhumbra, S.S., Paknikar, R.S., and Dressler, G.R. 2012. Epigenetic mechanisms of Groucho/Grg/TLE mediated transcriptional repression. *Mol Cell* **45**: 185-195.

Patterson, S.E., Bird, N.C., and Devoto, S.H. 2010. BMP regulation of myogenesis in zebrafish. *Dev Dyn* **239**: 806-817.

Pennacchio, L.A., Ahituv, N., Moses, A.M., Prabhakar, S., Nobrega, M.A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K.D., Plajzer-Frick, I., Akiyama, J., De Val, S., Afzal, V., Black, B.L., Couronne, O., Eisen, M.B., Visel, A., and Rubin, E.M. 2006. In vivo enhancer analysis of human conserved non-coding sequences. *Nature* **444**: 499-502.

Persson, M., Stamataki, D., te Welscher, P., Andersson, E., Bose, J., Ruther, U., Ericson, J., and Briscoe, J. 2002. Dorsal-ventral patterning of the spinal cord requires Gli3 transcriptional repressor activity. *Genes Dev* **16**: 2865-2878.

Phelan, S.A., Ito, M., and Loeken, M.R. 1997. Neural tube defects in embryos of diabetic mice: role of the Pax-3 gene and apoptosis. *Diabetes* **46**: 1189-1197.

Pick, L., Schier, A., Affolter, M., Schmidt-Glenewinkel, T., and Gehring, W.J. 1990. Analysis of the ftz upstream element: germ layer-specific enhancers are independently autoregulated. *Genes Dev* **4**: 1224-1239.

Pierani, A., Moran-Rivard, L., Sunshine, M.J., Littman, D.R., Goulding, M., and Jessell, T.M. 2001. Control of interneuron fate in the developing spinal cord by the progenitor homeodomain protein Dbx1. *Neuron* **29**: 367-384.

Postlethwait, J.H., Yan, Y.L., Gates, M.A., Horne, S., Amores, A., Brownlie, A., Donovan, A., Egan, E.S., Force, A., Gong, Z., Goutel, C., Fritz, A., Kelsh, R., Knapik, E., Liao, E., Paw, B., Ransom, D., Singer, A., Thomson, M., Abduljabbar, T.S., Yelick, P., Beier, D., Joly, J.S., Larhammar, D., Rosa, F., Westerfield, M., Zon, L.I., Johnson, S.L., and Talbot, W.S. 1998. Vertebrate genome evolution and the zebrafish gene map. *Nat Genet* **18**: 345-349.

Prasad, M.S. and Paulson, A.F. 2011. A combination of enhancer/silencer modules regulates spatially restricted expression of cadherin-7 in neural epithelium. *Dev Dyn* **240**: 1756-1768.

Pruitt, S.C. 1994. Primitive streak mesoderm-like cell lines expressing Pax-3 and Hox gene autoinducing activities. *Development* **120**: 37-47.

Pruitt, S.C., Bussman, A., Maslov, A.Y., Natoli, T.A., and Heinaman, R. 2004. Hox/Pbx and Brn binding sites mediate Pax3 expression in vitro and in vivo. *Gene Expr Patterns* **4**: 671-685.

Ptashne, M. 1986. Gene regulation by proteins acting nearby and at a distance. *Nature* **322**: 697-701.

Qiu, M., Shimamura, K., Sussel, L., Chen, S., and Rubenstein, J.L. 1998. Control of anteroposterior and dorsoventral domains of Nkx-6.1 gene expression relative to other Nkx genes during vertebrate CNS development. *Mech Dev* **72**: 77-88.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**: 279-283.

Ransick, A. and Davidson, E.H. 2006. cis-regulatory processing of Notch signaling input to the sea urchin glial cells missing gene during mesoderm specification. *Dev Biol* **297**: 587-602.

Ransick, A. and Davidson, E.H. 2012. Cis-regulatory logic driving glial cells missing: self-sustaining circuitry in later embryogenesis. *Dev Biol* **364**: 259-267.

Relaix, F., Polimeni, M., Rocancourt, D., Ponzetto, C., Schafer, B.W., and Buckingham, M. 2003. The transcriptional activator PAX3-FKHR rescues the defects of Pax3 mutant mice but induces a myogenic gain-of-function phenotype with ligand-independent activation of Met signaling in vivo. *Genes Dev* **17**: 2950-2965.

Relaix, F., Rocancourt, D., Mansouri, A., and Buckingham, M. 2004. Divergent functions of murine Pax3 and Pax7 in limb muscle development. *Genes Dev* **18**: 1088-1105.

Ribes, V., Balaskas, N., Sasai, N., Cruz, C., Dessaud, E., Cayuso, J., Tozer, S., Yang, L.L., Novitch, B., Marti, E., and Briscoe, J. 2010. Distinct Sonic Hedgehog signaling dynamics specify floor plate and ventral neuronal progenitors in the vertebrate neural tube. *Genes Dev* **24**: 1186-1200.

Ridgeway, A.G. and Skerjanc, I.S. 2001. Pax3 is essential for skeletal myogenesis and the expression of Six1 and Eya2. *J Biol Chem* **276**: 19033-19039.

Roberts, A. 2000. Early functional organization of spinal neurons in developing lower vertebrates. *Brain Res Bull* **53**: 585-593.

Robichaud, G.A., Nardini, M., Laflamme, M., Cuperlovic-Culf, M., and Ouellette, R.J. 2004. Human Pax-5 C-terminal isoforms possess distinct transactivation properties and are differentially modulated in normal and malignant B cells. *J Biol Chem* **279**: 49956-49963.

Robson, E.J., He, S.J., and Eccles, M.R. 2006. A PANorama of PAX genes in cancer and development. *Nat Rev Cancer* **6**: 52-62.

Ruiz i Altaba, A., Placzek, M., Baldassare, M., Dodd, J., and Jessell, T.M. 1995. Early stages of notochord and floor plate development in the chick embryo defined by normal and induced expression of HNF-3 beta. *Dev Biol* **170**: 299-313.

Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M., and Shiroishi, T. 2005. Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development* **132**: 797-803.

Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W.W., and Lenhard, B. 2004a. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res* **32**: D91-94.

Sandelin, A., Wasserman, W.W., and Lenhard, B. 2004b. ConSite: web-based prediction of regulatory elements using cross-species comparison. *Nucleic Acids Res* **32**: W249-252.

Sander, M., Paydar, S., Ericson, J., Briscoe, J., Berber, E., German, M., Jessell, T.M., and Rubenstein, J.L. 2000. Ventral neural patterning by Nkx homeobox genes: Nkx6.1 controls somatic motor neuron and ventral interneuron fates. *Genes Dev* **14**: 2134-2139.

Sasaki, H. and Hogan, B.L. 1994. HNF-3 beta as a regulator of floor plate development. *Cell* **76**: 103-115.

Sasaki, H. and Hogan, B.L. 1996. Enhancer analysis of the mouse HNF-3 beta gene: regulatory elements for node/notochord and floor plate are independent and consist of multiple sub-elements. *Genes Cells* **1**: 59-72.

Schier, A.F. and Gehring, W.J. 1992. Direct homeodomain-DNA interaction in the autoregulation of the fushi tarazu gene. *Nature* **356**: 804-807.

Schmitz, B., Papan,C.,and Campos-Ortega,J.A. 1993. Neurulation in the anterior trunk region of the zebrafish Brachydanio rerio. *Development Genes and Evolution* **202**: 250-259.

Seger, C., Hargrave, M., Wang, X., Chai, R.J., Elworthy, S., and Ingham, P.W. 2011. Analysis of Pax7 expressing myogenic cells in zebrafish muscle development, injury, and models of disease. *Dev Dyn* **240**: 2440-2451.

Serpente, P., Tumpel, S., Ghyselinck, N.B., Niederreither, K., Wiedemann, L.M., Dolle, P., Chambon, P., Krumlauf, R., and Gould, A.P. 2005. Direct crossregulation between retinoic acid receptor beta and Hox genes during hindbrain segmentation. *Development* **132**: 503-513.

Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., Weinstock, G.M., Wilson, R.K., Gibbs, R.A., Kent, W.J., Miller, W., and Haussler, D. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**: 1034-1050.

Smith, S.T. and Jaynes, J.B. 1996. A conserved region of engrailed, shared among all en-, gsc-, Nk1-, Nk2- and msh-class homeoproteins, mediates active transcriptional repression in vivo. *Development* **122**: 3141-3150.

Soleimani, V.D., Punch, V.G., Kawabe, Y., Jones, A.E., Palidwor, G.A., Porter, C.J., Cross, J.W., Carvajal, J.J., Kockx, C.E., van Ijcken, W.F., Perkins, T.J., Rigby, P.W., Grosveld, F., and Rudnicki, M.A. 2012. Transcriptional dominance of pax7 in adult myogenesis is due to high-affinity recognition of homeodomain motifs. *Dev Cell* **22**: 1208-1220.

Sosa-Pineda, B., Chowdhury, K., Torres, M., Oliver, G., and Gruss, P. 1997. The Pax4 gene is essential for differentiation of insulin-producing beta cells in the mammalian pancreas. *Nature* **386**: 399-402.

Spivakov, M. and Fisher, A.G. 2007. Epigenetic signatures of stem-cell identity. *Nat Rev Genet* **8**: 263-271.

Splinter, E. and de Laat, W. 2011. The complex transcription regulatory landscape of our genome: control in three dimensions. *EMBO J* **30**: 4345-4355.

Stamataki, D., Ulloa, F., Tsoni, S.V., Mynett, A., and Briscoe, J. 2005. A gradient of Gli activity mediates graded Sonic Hedgehog signaling in the neural tube. *Genes Dev* **19**: 626-641.

Stellabotte, F. and Devoto, S.H. 2007. The teleost dermomyotome. *Dev Dyn* **236**: 2432-2443.

Stern, C.D. 2005. Neural induction: old problem, new findings, yet more questions. *Development* **132**: 2007-2021.

Storey, K.G., Goriely, A., Sargent, C.M., Brown, J.M., Burns, H.D., Abud, H.M., and Heath, J.K. 1998. Early posterior neural tissue is induced by FGF in the chick embryo. *Development* **125**: 473-484.

Strahle, U., Blader, P., Henrique, D., and Ingham, P.W. 1993. Axial, a zebrafish gene expressed along the developing body axis, shows altered expression in cyclops mutant embryos. *Genes Dev* **7**: 1436-1446.

Takahashi, M. and Osumi, N. 2008. Expression study of cadherin7 and cadherin20 in the embryonic and adult rat central nervous system. *BMC Dev Biol* **8**: 87.

Tanabe, Y., William, C., and Jessell, T.M. 1998. Specification of motor neuron identity by the MNR2 homeodomain protein. *Cell* **95**: 67-80.

Thattaliyath, B.D., Firulli, B.A., and Firulli, A.B. 2002. The basic-helix-loop-helix transcription factor HAND2 directly regulates transcription of the atrial naturetic peptide gene. *J Mol Cell Cardiol* **34**: 1335-1344.

The ENCODE project consortium. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**: 799-816.

Thermes, V., Grabher, C., Ristoratore, F., Bourrat, F., Choulika, A., Wittbrodt, J., and Joly, J.S. 2002. I-SceI meganuclease mediates highly efficient transgenesis in fish. *Mech Dev* **118**: 91-98.

Timmer, J.R., Wang, C., and Niswander, L. 2002. BMP signaling patterns the dorsal and intermediate neural tube via regulation of homeobox and helix-loop-helix transcription factors. *Development* **129**: 2459-2472.

Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., and de Laat, W. 2002. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell* **10**: 1453-1465.

Treisman, J., Harris, E., and Desplan, C. 1991. The paired box encodes a second DNA-binding domain in the paired homeo domain protein. *Genes Dev* **5**: 594-604.

Tremblay, P., Pituello, F., and Gruss, P. 1996. Inhibition of floor plate differentiation by Pax3: evidence from ectopic expression in transgenic mice. *Development* **122**: 2555-2567.

Valenta, T., Gay, M., Steiner, S., Draganova, K., Zemke, M., Hoffmans, R., Cinelli, P., Aguet, M., Sommer, L., and Basler, K. 2011. Probing transcription-specific outputs of beta-catenin in vivo. *Genes Dev* **25**: 2631-2643.

Vallstedt, A., Muhr, J., Pattyn, A., Pierani, A., Mendelsohn, M., Sander, M., Jessell, T.M., and Ericson, J. 2001. Different levels of repressor activity assign redundant and specific roles to Nkx6 genes in motor neuron and interneuron specification. *Neuron* **31**: 743-755.

Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L.A. 2007. VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Res* **35**: D88-92.

Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., Afzal, V., Ren, B., Rubin, E.M., and Pennacchio, L.A. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**: 854-858.

Warming, S., Costantino, N., Court, D.L., Jenkins, N.A., and Copeland, N.G. 2005. Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Res* **33**: e36.

Whiting, J., Marshall, H., Cook, M., Krumlauf, R., Rigby, P.W., Stott, D., and Allemann, R.K. 1991. Multiple spatially specific enhancers are required to reconstruct the pattern of Hox-2.6 gene expression. *Genes Dev* **5**: 2048-2059.

Wijgerde, M., McMahon, J.A., Rule, M., and McMahon, A.P. 2002. A direct requirement for Hedgehog signaling for normal specification of all ventral progenitor domains in the presumptive mammalian spinal cord. *Genes Dev* **16**: 2849-2864.

Wilson, D., Sheng, G., Lecuit, T., Dostatni, N., and Desplan, C. 1993. Cooperative dimerization of paired class homeo domains on DNA. *Genes Dev* **7**: 2120-2134.

Wilson, S.W., Brand, M., and Eisen, J.S. 2002. Patterning the zebrafish central nervous system. *Results Probl Cell Differ* **40**: 181-215.

Woolfe, A., Goodson, M., Goode, D.K., Snell, P., McEwen, G.K., Vavouri, T., Smith, S.F., North, P., Callaway, H., Kelly, K., Walter, K., Abnizova, I., Gilks, W., Edwards, Y.J., Cooke, J.E., and Elgar, G. 2005. Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol* **3**: e7.

Woolfe, A. and Elgar, G. 2007. Comparative genomics using Fugu reveals insights into regulatory subfunctionalization. *Genome Biol* **8**: R53.

Woolfe, A., Goode, D.K., Cooke, J., Callaway, H., Smith, S., Snell, P., McEwen, G.K., and Elgar, G. 2007. CONDOR: a database resource of developmentally associated conserved non-coding elements. *BMC Dev Biol* **7**: 100.

Woolfe, A. and Elgar, G. 2008. Organization of conserved elements near key developmental regulators in vertebrate genomes. *Adv Genet* **61**: 307-338.

Xu, W., Rould, M.A., Jun, S., Desplan, C., and Pabo, C.O. 1995. Crystal structure of a paired domain-DNA complex at 2.5 A resolution reveals structural basis for Pax developmental mutations. *Cell* **80**: 639-650.

Xu, H.E., Rould, M.A., Xu, W., Epstein, J.A., Maas, R.L., and Pabo, C.O. 1999. Crystal structure of the human Pax6 paired domain-DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding. *Genes Dev* **13**: 1263-1275.

Zambrowicz, B.P., Imamoto, A., Fiering, S., Herzenberg, L.A., Kerr, W.G., and Soriano, P. 1997. Disruption of overlapping transcripts in the ROSA beta geo 26 gene trap strain leads to widespread expression of beta-galactosidase in mouse embryos and hematopoietic cells. *Proc Natl Acad Sci USA* **94**: 3789-3794.

Zechner, D., Muller, T., Wende, H., Walther, I., Taketo, M.M., Crenshaw, E.B., 3rd, Treier, M., Birchmeier, W., and Birchmeier, C. 2007. Bmp and Wnt/beta-catenin signals control expression of the transcription factor Olig3 and the specification of spinal cord neurons. *Dev Biol* **303**: 181-190.

Zeitlinger, J., Zinzen, R.P., Stark, A., Kellis, M., Zhang, H., Young, R.A., and Levine, M. 2007. Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the Drosophila embryo. *Genes Dev* **21**: 385-390.